Author/s:
Johnson, Christina Elizabeth

Title:
Guiding educators in learner-centred feedback: designing, testing and refining the Feedback Quality Instrument

Date:
2020

Persistent Link:
https://hdl.handle.net/11343/268099

Guiding educators in learner-centred feedback: designing, testing and refining the Feedback Quality Instrument

Dr Christina Elizabeth Johnson

orcid.org/0000-0002-4209-8419

December 2020

Department of Medical Education

School of Medicine

Faculty of Medicine, Dentistry and Health Sciences,

The University of Melbourne

Submitted with publication in total fulfilment of the degree of Doctor of Philosophy

Abstract

Feedback is important but problematic in the health workplace. The aim of this research is to promote quality feedback discussions that enhance learner outcomes. This thesis reports the process of developing, testing and refining an instrument to guide and systematically evaluate an educator's role in quality face-to-face feedback in clinical practice. It is underpinned by social constructivism, and uses a multiphase mixed methods design to achieve the research objective.

The research was conducted in three phases. Phase 1 comprised a systematic review and meta-analysis to summarise the empirical evidence available concerning the impact of face-to-face feedback on workplace task performance involving health professionals. Phase 2 focused on the development of a provisional feedback instrument. An extensive narrative literature review identified the distinct elements of an educator's role with the potential to influence learner outcomes. These elements were then operationalised into corresponding observable educator behaviours and refined in collaboration with an expert panel using a Delphi process. This resulted in consensus on a set of twenty-five descriptions of educator behaviours that manifest quality feedback, which constituted a provisional instrument. Phase 3 involved refining the provisional instrument to create the Feedback Quality Instrument using quantitative and qualitative analyses. Thirty-six videos of authentic clinical feedback discussions were collected. Six raters independently administered the provisional instrument to evaluate educator behaviours seen in the videos, compared with recommended behaviours in the provisional instrument, using a Likert scale: 0 = not seen, 1 = done somewhat, 2 = done consistently. This enabled usability testing and generated 174 sets of ratings data. Quantitative analysis of these ratings data were used to describe a) how commonly each recommended behaviour was demonstrated by the group of educators and b) variation across educators' practice, which was achieved by describing how many of the recommended behaviours each individual educator demonstrated. This afforded rare observational insights into contemporary feedback interactions, confirmed a gap between recommended and workplace practice, and indicated priorities for professional development initiatives. These ratings data were also used for

psychometric analysis of the provisional instrument using multifaceted Rasch model analysis and exploratory factor analysis. In addition, using thematic analysis, qualitative analysis of the video data was performed to explore two key emergent phenomena 1) the potential for promoting learners' own evaluative judgement and 2) ways in which educators could help to cultivate psychological safety with learners within feedback conversations. The qualitative analyses further clarified desirable educator approaches to feedback, enabling refinement of the instrument and addressing gaps identified during psychometric testing. Finally, the provisional instrument was refined, informed by the usability testing, psychometric testing, qualitative analysis, and foundational literature and theory. This resulted in the Feedback Quality Instrument (FQI), ready-for-use in clinical practice. The FQI includes 25 items, grouped into five domains: *set the scene*, *analyse performance*, *plan improvements*, *foster learner agency*, and *foster psychological safety*. These domains characterise five core concepts underpinning quality feedback identified by exploratory factor analysis. The items offer practical guidance for clinicians in fostering learner-centred feedback in collaboration with learners by describing behaviours that engage, motivate and enable a learner to improve. The instrument also provides a platform for future research investigating the impact of specific components of feedback on learner outcomes in the clinical workplace. This PhD program of research leading to the FQI includes four published articles, one article under submission and one published book chapter.

Declaration

This thesis with publications comprises only original work conducted by me as the principal researcher towards the degree of Doctor of Philosophy, except where indicated in the preface.

The contribution of my co-authors in jointly-authored work included in this thesis has been clearly acknowledged.

Due acknowledgement has been made in the text to all other material used.

Data analyses were performed by me, except where indicated in the Preface.

The thesis is fewer than the maximum word limit in length.

Dr Christina Johnson Date: 22 December 2020

Preface

This thesis contains original research in Chapters 1-7, conducted by me as the principal researcher, under the supervision of Professor Elizabeth Molloy, (Professor of Work Integrated Learning, Department of Medical Education, University of Melbourne), Professor Jennifer Keating (Emeritus Professor, Department of Physiotherapy, Monash University) and Professor Michelle Leech (Head of the Medical Course and Deputy Dean, Faculty Medicine Nursing and Health Sciences, Monash University).

This PhD was started in the Health Professions Education and Educational Research (HealthPEER) Unit, Faculty of Medicine, Nursing and Health Sciences, Monash University. Then when my primary supervisor Professor Elizabeth Molloy moved to the Department of Medical Education, School of Medicine, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, I enrolled at the University of Melbourne, and my PhD continued there.

This research was commenced during my Master's candidature, which was completed in July 2014. An initial literature review and one round of the Delphi technique was included as a minor research project contributing to my Master's by coursework and research. This early introduction to key research methods provided a useful foundation for the PhD. In the PhD, which commenced in March 2015, the research conducted during the period of Master's candidature was extended and thoroughly revised, including additional data collection, data analysis and synthesis, and refinement to the items that eventually made up the provisional feedback instrument. The PhD research built on the Master's platform with two further rounds of the Delphi process and a face-to-face meeting of the expert panel that led to substantial revisions to the item set. In addition, the PhD candidature included continuing expansion and updating of the narrative literature review commenced during the period of Master's candidature to incorporate recent publications. This included publishing the research in the article presented in Chapter 3 of this thesis, published in BMC Medical Education in March 2016: Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement.

I gratefully acknowledge a grant towards the costs associated with the final face-to-face meeting of a Delphi expert panel from the Office of Learning and Teaching (OLT) and Higher Education and Research Development Society in Australasia (HERSDA) Researching New Directions in Learning and Teaching: Seed Funding to Support Project Mentoring and Collaboration.

Both quantitative and qualitative data analyses were predominantly performed by me. Further details are included in the 'Background' section prior to relevant chapters: qualitative analyses in Chapters 5 and 7, and quantitative analyses in Chapters 4 and 6.

This thesis includes four articles published in peer review journals, one article under submission and one published book chapter. My contribution and the contributions of my co-authors for each publication are detailed in the table below.

Contribution of authors to publications in the PhD

Thesis chapter	Publication title	Publication status	Nature and % of student contribution	Co-author name(s) and % of co-author's contribution
2	Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis.	Article published in BMJ Open	Led the study including study design; creating and conducting the literature search, study screening and quality assessment, data extraction, analysis and interpretation, and preparing the manuscript for publication (90%)	Jennifer Keating contributed to study design; study screening and quality assessment, data extraction, analysis and interpretation and assisted in the drafting of the manuscript; 5% Mihiri Weerasuria contributed to study screening and quality assessment, data extraction, and suggested revisions to the manuscript; 5%
3	Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement	Article published in BMC Medical Education	Led the study including study design; conducting the literature search, data extraction and synthesis; development of elements; development of initial behaviour statements and revisions during the Delphi process, and preparing the manuscript for publication (81%)	Elizabeth Molloy contributed to study design; synthesis of data from the literature; development of elements; development of initial behaviour statements and revisions during the Delphi process, and preparing the manuscript for publication design; 5% Jennifer Keating contributed to study design; synthesis of data from the literature; development of elements; development of initial behaviour statements and revisions during the Delphi process, and preparing the manuscript for publication design; 5% David Boud participated in the Delphi panel and suggested revisions to the manuscript; 1% Megan Dalton participated in the Delphi panel and suggested revisions to the manuscript; 1% Debra Kiegaldie participated in the Delphi panel and suggested revisions to the manuscript; 1% Margaret Hay participated in the Delphi panel and suggested revisions to the manuscript; 1%

Thesis chapter	Publication title	Publication status	Nature and % of student contribution	Co-author name(s) and % of co-author's contribution
				Barry McGrath participated in the Delphi panel and suggested revisions to the manuscript; 1% Wendy McKenzie participated in the Delphi panel and suggested revisions to the manuscript; 1% Kichu Nair participated in the Delphi panel and suggested revisions to the manuscript; 1% Debra Nestel participated in the Delphi panel and suggested revisions to the manuscript; 1% Claire Palermo participated in the Delphi panel and suggested revisions to the manuscript; 1%
4	Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis	Article published in BMC Medical Education	Led the study including study design; collected videos; video analysis; data analysis and interpretation; and preparing the manuscript for publication (84%)	Elizabeth Molloy contributed to study design; video analysis; data interpretation; and preparing the manuscript for publication; 5% Jennifer Keating contributed to study design; video analysis; data analysis and interpretation; and preparing the manuscript for publication; 5% Michelle Leech contributed to video analysis and preparing the manuscript for publication; 2% Melanie Farlie contributed to video analysis and suggested revisions to the manuscript; 2% Fiona Kent contributed to video analysis and suggested revisions to the manuscript; 2%

Thesis chapter	Publication title	Publication status	Nature and % of student contribution	Co-author name(s) and % of co-author's contribution
5	Building evaluative judgement through the process of feedback	Book chapter published in Developing evaluative judgement in higher education: assessment for knowing and producing quality work	Led the study, including study design; qualitative analysis of video data; and preparing the manuscript for publication; 90%	Elizabeth Molloy contributed to study design; qualitative analysis of video data; data interpretation; and preparing the manuscript for publication; 10%
6	Psychological safety in feedback: What does it look like and how can educators work with learners to foster it?	Article published in Medical Education	Led the study including study design; qualitative analysis of video data; and preparing the manuscript for publication; 90%	Elizabeth Molloy contributed to study design; qualitative analysis of video data; and preparing the manuscript for publication; 5% Jennifer Keating contributed to study design; qualitative analysis of video data; and preparing the manuscript for publication; 5%
7	The Feedback Quality Instrument: A guide for health professional educators in fostering learner-centred discussions	Submitted to Medical Education	Led the study including study design; data analysis and data interpretation; development of the Feedback Quality Instrument and preparing the manuscript for publication; 85%	Elizabeth Molloy contributed to study design; development of the Feedback Quality Instrument and preparing the manuscript for publication; 5% Jennifer Keating contributed to study design; data analysis and interpretation; development of the Feedback Quality Instrument and preparing the manuscript for publication; 5% Michelle Leech contributed to prior video analysis using provisional instrument and preparing the manuscript for publication; 2% Peter Congdon conducted Rasch analysis and assisted with interpretation; and suggested revisions to the manuscript; 1% Melanie Farlie contributed to prior video analysis using provisional instrument and suggested revisions to the manuscript; 1% Fiona Kent contributed to prior video analysis using provisional instrument and suggested revisions to the manuscript; 1%

Publications

Published

1. Johnson CE, Keating JL, Molloy EK. Psychological safety in feedback: What does it look like and how can educators work with learners to foster it? Medical Education. 2020;54(6):559-570. (Chapter 6)

JCR Impact factor 4.570 (2019); Q1 for education; 2 citations

2. Johnson CE, Weerasuria MP, Keating JL. Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis. *BMJ Open.* 2020;10(3):e030672 (Chapter 2)

JCR impact factor 2.96 (2019); Q2 for medicine; 1 citation

3. Johnson CE, Keating JL, Farlie MK, Kent F, Leech M, Molloy EK. Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. BMC Medical Education. 2019;19(1):129. (Chapter 4)

JCR Impact factor 1.831; Q2 for education; 0 citations

4. Johnson CE, Molloy EK. Building evaluative judgement through the process of feedback. In: Boud D, Ajjawi R, Dawson P, Tai J, eds. Developing evaluative judgement in higher education. Assessment for knowing and producing quality work. London: Routledge; 2018:166-175. ISBN 9781138089358 (Chapter 5)

5. Johnson CE, Keating JL, Boud DJ, Dalton M, Kiegaldie D, Hay M, McGrath B, McKenzie WA, Nair KBR, Nestel D, Palermo C, Molloy EK. Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement. BMC Medical Education. 2016;16(1):96. (Chapter 3)

JCR Impact factor 1.572; Q2 for education; 33 citations

Submitted

Johnson CE, Keating JL, Leech M, Congdon P, Kent F, Farlie MK, Molloy EK. The Feedback Quality Instrument: A guide for health professional educators in fostering learner-centred discussions. Submitted for publication to Medical Education, December 2020. (Chapter 7)

Podcast

Johnson CE, Eva K. Psychological safety in feedback: What does it look like and how can educators work with learners to foster it? This paper was selected by the Editor-in-Chief, Professor Kevin Eva for the Medical Education podcast interview in June 2020. Link here

Prize

Melbourne Medical School Annual Publication Prize (students) 2020, University of Melbourne.

This prize was awarded to Dr Christina Johnson for the following research article, described in Chapter 6.

Johnson CE, Keating JL, Molloy EK. Psychological safety in feedback: What does it look like and how can educators work with learners to foster it? Medical Education. 2020;54(6):559-570. https://doi.org/10.1111/medu.14154

The Publication Prize aims to recognise and reward high quality research produced by graduate research students in the Melbourne Medical School and encourage them to further disseminate their research findings through publications and presentations. The broad intention of the Prize is to recognise all forms of research that take place across the School, including basic, translational and implementation research.

Conference Presentations

Johnson CE, Weerasuria MP, Keating JL. 'A systematic review: the impact of feedback on health professionals' task performance – what's the evidence?' International Association for Medical Education (AMEE) annual conference (virtual conference due to COVID), September 2020.

Johnson CE, Keating JL, Molloy EK. The impact of psychological safety on dialogue during feedback. Accepted for Australian and New Zealand Association for Health Professional Educators (ANZAHPE) annual conference (conference cancelled due to COVID), Melbourne, July 2020.

Johnson CE, Keating JK, Kent F, Farlie M, Leech M, Molloy EM. How closely does educator behaviour during feedback in contemporary clinical practice align with published recommendations: an observational study of 36 authentic formal feedback episodes across the health professions. International Association for Medical Education (AMEE) annual conference, Basel, Switzerland, August 2018.

Johnson CE, Molloy EM. The place of evaluative judgement in verbal feedback sessions: an observational study across the health professions. International Association for Medical Education (AMEE) annual conference, Helsinki, August 2017.

Johnson CE, Keating J, Boud D, Dalton M, Kiegaldie D, Hay M, McGrath B, McKenzie W, Nair K, Nestel D, Palermo C, Molloy E. Feedback Quality Instrument (FQI): Development of a verbal feedback quality assessment instrument for health professionals in the workplace.

International Association for Medical Education (AMEE) annual conference, Milan, September 2014.

Other publications and presentations arising during candidature

(related to, but not directly arising from, thesis work)

Publication

Johnson CE, Watling CJ, Keating JL, Molloy EK. Effective feedback conversations in clinical practice. In: Nestel D, Reedy G, McKenna L, Gough S, eds. Clinical Education for the Health Professions: Theory and Practice. Singapore: Springer; *In press* (*accepted September 2020).* (Appendix 5)

Presentations

Johnson CE. Light bulb moments for me while investigating feedback, and Psychological safety. Faculty: Foundations of Health Professions Education unit, Graduate Certificate in Clinical Simulation, Monash University, September 2020.

Johnson CE. Feedback in Clinical Practice. Invited speaker: Graduate Certificate in Health Professions Education, Monash University, September 2020.

Johnson CE. Feedback in clinical practice: what works? Invited speaker: Graduate Certificate in Clinical Teaching, Excellence in Clinical Education (EXCITE) program, The University of Melbourne April 2020.

Johnson CE. Feedback conversations in clinical practice. Faculty: Learning Conversations: Feedback and debriefing practices in clinical practice and in simulation (short course), Monash University, March 2020.

Johnson CE. The role and purpose of feedback in clinical practice - the current theory and practices, and Delivering feedback in clinical practice - the models which assist your clinical practice. Conference workshop: World Federation for Ultrasound in Medicine and Biology (WFUMB) annual conference, Melbourne, September 2019.

Johnson CE. Feedback for learning. Faculty: Foundations of Health Professions Education, Graduate Certificate in Clinical Simulation / Health Professions Education, Monash University, September 2019.

Johnson CE. Effective Feedback: practical tips for educators. Conference workshop: Australian and New Zealand Prevocational Medical Education Forum, Melbourne, November 2018.

Johnson CE. Effective Feedback. Invited speaker: Graduate Certificate in Clinical Teaching, Excellence in Clinical Education (EXCITE) program, The University of Melbourne September 2017.

Molloy EM, Johnson CE. The Problems and Prospects of Feedback in Workplace-Based Learning. Keynote address: Postgraduate Medical Council of Victoria (PMCV) Annual Symposium, Melbourne, May 2017.

Acknowledgements

Thank you to my supervisors Professor Elizabeth Molloy, Professor Jennifer Keating and Professor Michelle Leech. I have benefited enormously from your expertise, guidance and support along this journey, for which I am deeply grateful.

Thank you to all the experts on the Delphi panel. Your diverse perspectives and useful insights greatly enhanced the instrument.

Thank you to the educators and learners who volunteered to video their feedback conversations. I am indebted to you for taking this professional risk and partnering with us to advance our knowledge about 'what quality feedback discussions' could look like.

Table of Contents

1. Johnson CE, Weerasuria MP, Keating JL. Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis. BMJ Open. 2020;10(3):e030672 (Chapter 2)
2. Johnson CE, Keating JL, Boud DJ, Dalton M, Kiegaldie D, Hay M, McGrath B, McKenzie WA, Nair KBR, Nestel D, Palermo C, Molloy EK. Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement. BMC Medical Education. 2016;16(1):96. (Chapter 3)
3. Johnson CE, Keating JL, Farlie MK, Kent F, Leech M, Molloy EK. Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. BMC Medical Education. 2019;19(1):129. (Chapter 4)
4. Johnson CE, Keating JL, Molloy EK. Psychological safety in feedback: What does it look like and how can educators work with learners to foster it? Medical Education. 2020;54(6):559-570. (Chapter 6)
5. Johnson CE, Watling CJ, Keating JL, Molloy EK. Effective feedback conversations in clinical practice. In: Nestel D, Reedy G, McKenna L, Gough S, eds. Clinical Education for the Health Professions: Theory and Practice. Singapore: Springer; In press (accepted September 2020).

List of Tables

List of Figures

Copyright material

Chapter 1

Introduction

1.1 Summary

This 'PhD with publication' describes the process of developing, testing and refining the Feedback Quality Instrument, designed to guide and evaluate an educator's role during face-to-face feedback in clinical practice. A multi-phase mixed methods design was employed, conducted in three phases. Phase 1 investigated the current evidence regarding the effects of face-to-face verbal feedback on health professionals' performance via a systematic review. Phase 2 involved the development of a provisional feedback instrument. A narrative literature review into best-practice verbal feedback was conducted followed by a Delphi process to develop expert consensus on a set of educator behaviours likely to be associated with quality verbal feedback, essentially forming a provisional feedback instrument. Phase 3 focused on revising the provisional instrument to create the Feedback Quality Instrument, using video-recordings of authentic feedback interactions in clinical practice. This process involved both quantitative analyses, based on ratings generated by using the provisional instrument to analyse educators' practice, and qualitative analyses exploring evaluative judgement and psychological safety in the feedback videos. Refinements to the provisional instrument were informed by usability testing, psychometric testing and qualitative analysis to produce the current version of the Feedback Quality Instrument.

This chapter opens by considering the role of feedback in healthcare. Following this a summary of the PhD is presented, including the knowledge gaps, research aims and research questions, and chapter outlines. Lastly, a personal perspective and my epistemic stance are presented.

1.2 Background

Across many disciplines, experts agree that feedback can have a powerful influence on learning and performance.[1-6]

Learners want feedback too, as they believe it can help them to improve their skills.[7-9] Yet studies of feedback in practice reveal multiple concerns related to both learners and educators, as well as feedback outcomes.[4,10-13] Widespread dissatisfaction with feedback has resulted in universities and training institutions pouring resources into trying to improve 'the feedback problem'.[10,14-16]

The concept of feedback emerged from biology and cybernetics. In these disciplines, feedback refers to a system that monitors important parameters and uses the data to adjust its output and optimise performance.[6] One example in the human body is the kidney: when blood sodium levels are high, the kidney excretes more sodium in the urine, until the blood sodium level reaches satisfactory levels.[17] This is an ongoing process, continually using performance information to drive desired results.

In the context of enhancing a person's performance of a task, definitions of feedback reflect changing theoretical perspectives over time. Ende, in his seminal 1983 article that focused on junior doctors in clinical practice, defined feedback as "an informed, non-evaluative, objective appraisal of performance intended to improve clinical skills."[3(p779)] Ende highlighted the importance of educators watching learners as they performed clinical tasks, to provide the basis for their feedback comments. A key innovation in this definition was the shift in focus to a formative perspective with the aim of helping learners to improve, from a summative one, that is, judging whether performance met expected standards or not. However, this definition centres on feedback as 'performance information delivered by an educator', with no mention of the learner's role.

This narrow focus solely on 'the delivery of information' was challenged by Sadler, working in higher education. He turned attention to the *effect* of performance information and whether or not it resulted in improved performance. This aligned more closely with the original concept of biological feedback loops. Sadler supported a definition by Ramaprasad that feedback "is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way."[18(p4)] In the field of education, Hattie and Timperley later defined feedback

as “information provided by an agent (teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding.”[1(p81)] Here the procedure is viewed from the learner’s perspective but remains characterised as ‘information provision’. It recognises the learner’s own analysis and experience as key contributors, and broadens the valuable effects beyond performance, to include understanding.

Spanning both higher education and health professions education, in 2013 Boud and Molloy proposed a new definition, grounded in constructivism. They defined feedback as “a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, in order to generate improved work.”[19(p6)] Here again, learners are the focal point; they are positioned as active participants seeking and using performance information to improve. Constructivism asserts that people learn by actively creating mental models, which help them make sense of the world. When an event occurs, if their existing knowledge is insufficient to explain the experience, this drives them to pursue greater understanding.[20] Social constructivism emphasises how knowledge is co-constructed during interactions between people, for example as they explain, justify, query or challenge ideas. Additional important features of Boud and Molloy’s definition include presenting feedback as a process, as opposed to a one-off event, and endorsing the principle that feedback information should have an identifiable impact on future performance. They argue that if courses were designed to include a set of nested tasks, requiring increasingly difficult but partly overlapping skills, then a learner could apply feedback information gained from one task to a subsequent task. This would enable the learner to see if the new strategy worked as planned and resulted in better performance, or whether they needed to revise the strategy in some way. This focus on tracing the effects of feedback information also enables educators to monitor and adjust their own approaches to feedback discussions. Boud and Molloy also argue that valuable performance information can come from multiple sources, including peer learners, work colleagues or consumers, in addition to educators. These different stakeholders have varied perspectives or knowledge related to specific aspects of performance, thereby adding important data for learners to consider when appraising their own performance.

Feedback in clinical practice

Providing quality healthcare requires health professionals to develop multiple competencies. Even after gaining a professional qualification, all clinicians are required to continue enhancing their skills. This may involve formal additional specialist training or continuous professional development, which is a condition for ongoing registration as a healthcare practitioner. These rigorous requirements reflect the complexity and ongoing evolution of modern healthcare. Typically, learning is embedded within clinical practice and occurs 'on the job' alongside more experienced health professionals during health care delivery.[21,22]

In workplace learning, feedback plays a central role in enabling learners to develop their expertise.[5,23] The aim of feedback is to enable a learner to improve by comparing their performance with the desired performance and considering ways to improve it.[3,19] Ideally, this will enable a learner to develop new insights and improvement strategies that they can use to enhance subsequent performance. Here the term 'performance' is used broadly to encompass attitude, knowledge, skill or learning strategy. Common face-to-face feedback scenarios include 1) brief, spontaneous, off-the-record exchanges during a specific activity sometimes referred to as 'informal feedback' or 'feedback on the run'[24]; 2) workplace-based assessments, involving a feedback discussion related to specific performance of a clinical task, for example a practical procedure or clinical interaction[4]; or 3) a more comprehensive summary review of clinical performance at the end of a day or mid/end of a clinical attachment. Generally, feedback conversations involve a 'learner', a clinician or student whose work performance is being reviewed, and an 'educator', typically a senior clinician in the same discipline. Alternatively, the educator may be a peer or a more junior colleague in the same discipline or a clinician from a different health profession.

Other feedback modalities that occur in clinical practice include written or automated feedback. Written feedback alone is rare but workplace-based assessments and mid/end of attachment feedback interactions often incorporate a written record. An important benefit of written feedback is the option for the learner or educator to review

the original comments whenever desired, after obtaining the report.[25] Automated feedback is becoming increasingly prevalent in this digital age. Common examples include automated answers to questions within online learning modules; performance data generated by surgical simulators; or mannequins that beep during cardiopulmonary resuscitation when chest compression occurs to the desired depth. There are several advantages to automated feedback. Once a system is designed, it can run repeatedly at any hour; this expands training opportunities and saves human labour. The impersonal nature avoids any antagonism that may be caused by social interactions but also forgoes the potential benefits. As feedback comments are pre-set, they cannot be tailored for an individual nor extend beyond what is programmed, for example to answer additional queries or explain using an alternative approach.

Given the different types of 'feedback information' available to learners in a healthcare workplace, there is an increasing focus on the notion of learner 'feedback literacy'.[26-28] Carless and Boud defined feedback literacy as "the understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies.[26(p1316)] That is, focusing on how they seek information available from multiple sources, understand the implications for their particular context, and then apply it, in order to maximise the benefit of feedback practices.

Quality verbal face-to-face feedback

The research in this PhD will focus on scheduled verbal face-to-face feedback encounters following observation of learners' performance in the healthcare workplace. The unique advantage of face-to-face feedback is the opportunity for direct and immediate interaction. This allows each person to consider and respond to the other's contribution, for example to seek clarification, challenge or propose an alternative strategy. Such collaboration, as ideas are shared, explored and modified together, provides the basis for co-constructing understandings and improvement strategies, consistent with a social constructivism perspective.[20,23] In addition, the human interaction offers the potential to increase feelings of personal connection, inclusion, motivation and support.[29-32] Therefore feedback conversations provide the chance to

position learners so they feel empowered and psychologically safe to seek out and make the most of the many learning opportunities available during the interaction and more broadly within the healthcare workplace.

However there is little evidence to support the benefits or to advise regarding which constituents make such interactions successful in promoting learning.

Problems with feedback in practice

Despite the theoretical promise of feedback exchanges, they seem to be problematic in reality. Remarkably, a systematic review and analysis investigating the impact of feedback across many non-health disciplines reported that it resulted in deterioration in performance in approximately one third of cases.[2] In clinical practice, research into the influence of audit and feedback revealed that the effect varies, and that it does not always improve performance.[33,34] Feedback interactions may even cause undesirable outcomes including reduced motivation or negative emotions such as anger, humiliation or sadness.[35,36] In addition, learners frequently report they do not find feedback information credible, relevant, practical or easy to understand.[7,11,13,37] As for educators, they commonly do not feel confident in their feedback skills, dislike raising issues about substandard performance and worry about damaging collegial relationships with learners.[3,38] One reason for unpredictable outcomes may be that feedback interactions incorporate multiple components that have different effects. Currently there is no established consensus on beneficial constituents of feedback or how participants could foster high quality feedback interactions.[27,39]

1.3 Problem statement and knowledge gaps

The problem addressed in this thesis is that face-to-face feedback discussions in clinical practice have the potential to offer important benefits for learners but in practice, the effect is variable and may even cause harm. At the start of this research, key knowledge gaps regarding face-to-face feedback in health professions education included:

- empirical evidence regarding the impact of verbal feedback and an estimate of its magnitude

- consensus on key constituents in quality learner-centred feedback with the potential to result in beneficial learner outcomes
- clear descriptions of educator behaviours that promote quality learner-centred feedback
- clarity on characteristics of contemporary feedback practices and the potential learning needs of educators relating to feedback
- a method to systematically analyse the constituents of verbal feedback interactions in the clinical workplace

1.4 Research aim

The research aim was to create a comprehensive set of items that clearly articulate practical ways for educators to collaborate with learners, to promote high quality, learner-centred feedback. Such items could be used by educators and learners to help guide professional development at every level of training. A comprehensive item set (instrument) would enable analysis of educators' feedback practice, to prioritise areas for professional development. It could also provide the foundation for a subsequent research program to systematically investigate the influence of specific educator behaviours on subsequent learner outcomes.

1.5 Research questions

The research questions addressed in this thesis were as follows:

RQ 1: What is the current evidence available regarding the impact of face-to-face feedback on workplace task performance involving health professionals?

RQ 2: What are the distinct educator roles in quality feedback discussions that result in beneficial learner outcomes?

RQ 3: Could consensus be achieved on a comprehensive set of statements that describe important educator behaviours in quality feedback discussions?

RQ 4: What can be learnt about current feedback practices, by analysing data gained through administering the provisional instrument to systematically analyse feedback interactions in clinical practice?

RQ 5: What novel insights can be gained from thematic analysis of feedback episodes in clinical practice regarding emerging and under-researched arguments regarding best practice in feedback using the examples of i) learner evaluative judgement and ii) learner psychological safety?

RQ 6: Could a Feedback Quality Instrument be presented ready-for-use, by refining the provisional instrument based on multiple inputs including i) usability testing ii) psychometric analysis involving multifaceted Rasch model analysis and exploratory factor analysis iii) thematic analysis in under-researched areas?

1.6 Thesis Outline

This section will describe how each study contributed to developing the Feedback Quality Instrument. The research is reported in the following three phases, with the corresponding research questions and chapters in brackets (see Figure 1.1). Phase 1 clarified the available evidence and involved a systematic review and meta-analysis to investigate the impact of face-to-face feedback on workplace task performance involving health professionals (RQ1; Chapter 2). Phase 2 had two stages resulting in a comprehensive set of descriptions of educator behaviours designed to optimise learner outcomes, constituting a provisional feedback instrument. In Phase 2, Stage 1 an additional literature review was conducted to identify discrete elements of an educator's role in feedback discussions considered to enhance learner outcomes, supported by research information (RQ2; Chapter 3). Phase 2, Stage 2 involved operationalising these elements and using three rounds of a Delphi process plus a final face-to-face meeting with an expert panel to gain consensus on a set of descriptions of observable educator behaviours that exemplified qualify feedback (RQ3; Chapter 3). Phase 3 incorporated three stages resulting in the Feedback Quality Instrument, ready-to-use. Phase 3, Stage 1 started with collecting video-recordings of authentic feedback interactions of diverse health professionals in clinical practice. A team of raters undertook usability testing and independently administered the provisional instrument to systematically analyse educator's feedback practice seen in the feedback videos. First the ratings data were used to create an observational report on the characteristics of contemporary feedback practice (RQ4; Chapter 4). This information provided impetus and clarified priorities for guidance in quality feedback practice. Phase 3, Stage 2 involved parallel qualitative and additional quantitative analyses of the feedback video data to inform revisions to the provisional instrument. Qualitative analysis focused on exploring two important aspects in feedback, learner evaluative judgement and learner psychological safety using thematic analysis of the feedback video data (RQ5; evaluative judgement in Chapter 5 and psychological safety in chapter 6). These concepts were emerging in the literature but typical educator behaviours that could promote them, required for the feedback instrument, had not yet been well characterised. This research provided insights into approaches that educators had used to collaborate with learners

to cultivate psychological safety and evaluative judgement, to assist with item development. In parallel, psychometric analysis of the provisional instrument was conducted, involving further quantitative analyses of the video ratings data using multifaceted Rasch model analysis and exploratory factor analysis (RQ6; Chapter 7). Phase 3, Stage 3 culminated in the Feedback Quality Instrument based on refinements to the provisional instrument informed by instrument usability testing, psychometric analysis, thematic analysis and underlying research and theory (RQ6; Chapter 7). Finally, Chapter 8 wraps up the thesis by summarising the key findings that address each research question and describes the implications, future directions, and strengths and limitations of the research.

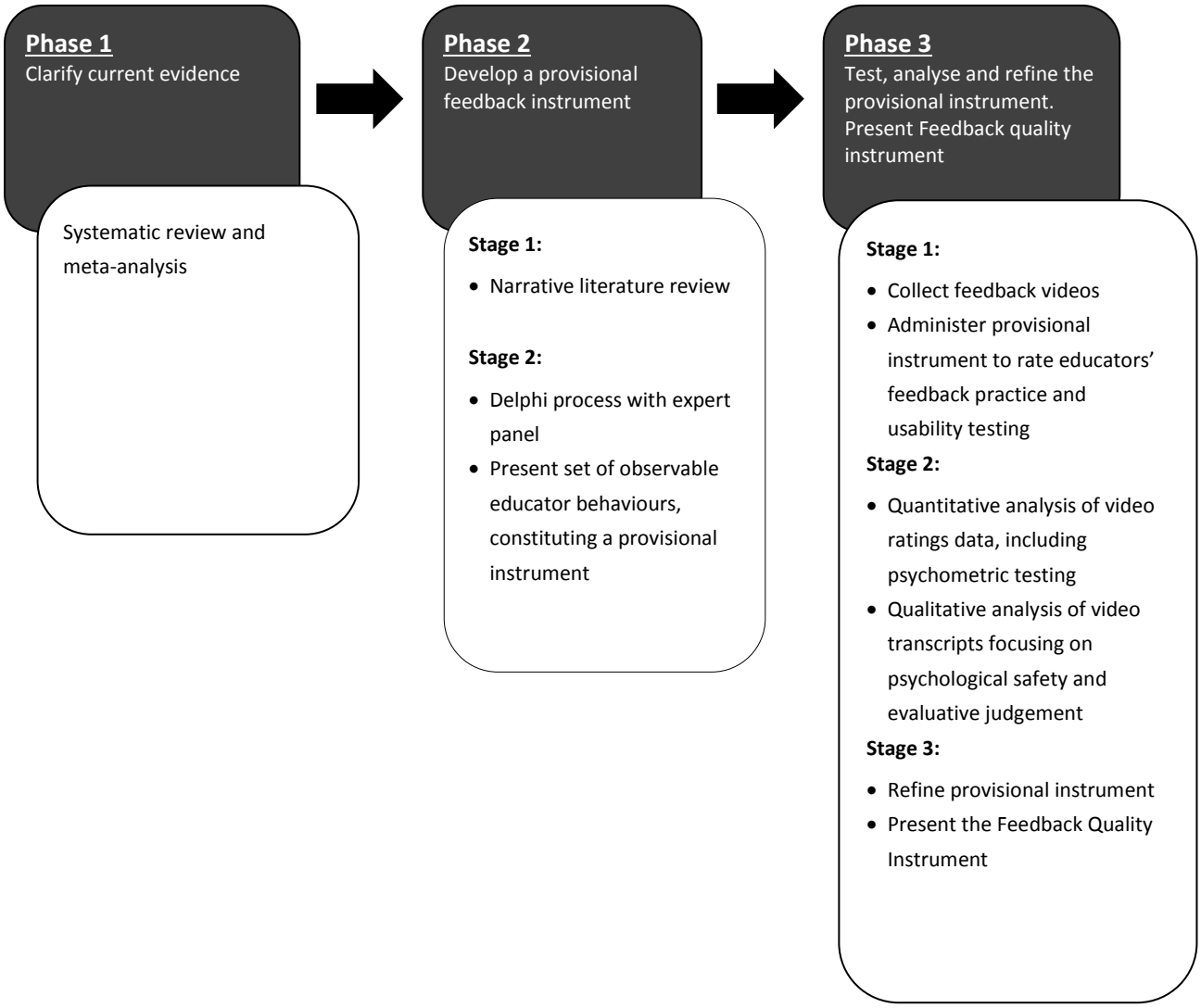


Figure 1.1: Overview of the three phases in the PhD research program

Personal perspective and epistemic stance

This thesis employed a mixed methods research approach, using a combination of quantitative and qualitative designs to address the different research questions.[40] Bazeley has called this a pragmatic approach.[41] This accommodates different traditions with their associated worldviews, in place of a singular approach, for the purposes of best answering the research questions. It has offered the chance for me to take a deep dive across contrasting epistemologies, with associated methods for capturing and making sense of data.

I trained, and still work, as a physician in public hospital practice. When I started formal study in health professions education, I already had a role as a regional supervisor for doctors undertaking specialist training in Geriatric Medicine. Feedback discussions were a core component of training for these junior doctors yet there seemed to be little information or training regarding 'how to do it well'. As I read more about feedback, two key questions began to emerge in my mind: 1) what are the important components in quality feedback discussions that optimise learner outcomes, and what is the supporting evidence? and 2) what can I do, as the educator, to foster quality feedback discussions? It was the desire to find answers to these questions that instigated this research.

As a medical doctor, I was immersed in the principles of 'evidence-based medicine'. This largely reflects a positivist approach, but I was not aware of this at the time. Positivism asserts that knowledge is obtained and verified by observing natural events using our senses, measuring objective data and quantifying effects using statistical analyses. Experiments are designed to test a hypothesis, analysis is deductive and results are generalised to similar settings. The researcher's role is simply to observe and analyse these events; they wish to be detached and to have no influence on them. The study design, variables of interest and analyses are all set beforehand and are not altered once the study starts.[42] The systematic review (Chapter 2) and psychometric analysis of the provisional feedback instrument (Chapter 7) in this thesis, in particular, align with this philosophy. The work in these two chapters has deepened and extended my quantitative research skills.

I completed a Graduate Certificate and then a Master's in Health Professions Education. With my strong scientific training, it took a while for me to appreciate alternative perspectives on knowledge, that is, different epistemologies. Gradually, my understanding of an interpretivist approach and the characteristics of rigorous qualitative research increased. In qualitative research approaches, findings emerge and influence the direction of the research. The research design is more open-ended and dynamic, so the methodological approach may change during the study if findings suggest alternative techniques or theorical lenses might be more useful in answering the research question. The researcher is intrinsically involved in the study; they influence the design and the findings as they interpret the participants' actions or responses. Therefore, it is important for the researcher to consider their own background and how they influence the research and visa versa; such 'researcher reflexivity' lays out the researcher's experiences and epistemic stance/s that are likely to influence the research process.[42] The two qualitative research articles exploring evaluative judgement (Chapter 5) and psychological safety (Chapter 6) align with this interpretivist approach. Conducting these qualitative analyses, has enabled me to cultivate and hone my qualitative research skills.

For the research portion of my Master's degree, I began researching 'effective feedback'. I then decided to take the next step and tackle a PhD with a focused exploration of what educators can do to cultivate learner-centred feedback interactions. During this time, I also changed work roles to become the director of medical education at the hospital, with primary responsibility for doctors in their first few years of training, and moved my major clinical role from geriatric medicine to acute general medicine. This combination of 'hands on' research and clinical practice with a major role in junior doctor training, has provided wonderful opportunities for 'theory' and 'practice' to inform and challenge each other. In addition, throughout this PhD, there have been many rich discussions with my supervisors. By using a mixed methods approach and tailoring the methodology according to the research question, I believe I have experienced 'the best of both worlds'.

This research focuses on verbal feedback conversations in health professions education, aligned with Boud and Molloy's[14p6] definition of feedback and underpinned by a theoretical framework of social constructivism, based on the work of Vygotsky.[20,43,44] From a social constructivist perspective, knowledge is actively constructed as people interact in a particular setting. In the context for my research, one example, is when a junior doctor discusses a patient's care with a senior doctor in a teaching hospital. In this apprenticeship model, the junior doctor learns how to perform necessary tasks for improving a patient's health, supported by a more experienced doctor. Learning complex skills in an authentic environment enables the junior doctor to appreciate the purpose of the task and how it contributes towards the overall process, thereby promoting understanding and motivation. Training in clinical settings also embeds the junior doctor in the cultural discourse of medical practice. This enables them to progressively internalise the 'way things are done around here'. Vygotsky described 'scaffolding', where the more senior doctor initially provides a lot of support, to enable the junior doctor to provide quality care to patients. Over time the junior doctor gains expertise, as they try to problem solve issues under the guidance of the senior doctor. Correspondingly, the senior doctor gradually adjusts and reduces their input, allowing the junior doctor to work more independently. Vygotsky also developed the concept of the 'zone of proximal development', referring to what a junior doctor could accomplish with assistance but not alone and indicating the next stage in their learning trajectory. All these aspects of social constructivism contribute to understanding effective ways to foster learning within feedback conversations. As knowledge is co-constructed, rather than handed down, or transmitted, this means that learners have a number of active roles to play in feedback interactions, other than 'being the attentive listener'. As learning is positioned as relational in this paradigm, it also means that both parties within a feedback conversation have things to lose and gain depending on the power dynamic, the stakes of the conversation (for example, a summative assessment, or otherwise) and the point in the learner's career trajectory. My stance about what constitutes learning and feedback has been influenced by the literature I have immersed myself in, the supervisors I have sought, and the questions that have guided this research. Notably, the research has also influenced my own understandings of what

learning means, and how I go about feedback in my workplace. While I have lots to say in reporting the findings in this PhD, during feedback encounters, I hope I say less, listen more and collaborate more effectively.

Chapter 2:

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

2.1 Chapter 2 Introduction

This PhD research was conducted in three phases, as outlined in the Thesis Outline (Section 1.6) within the Introduction (Chapter 1). Phase 1 involved clarifying the empirical evidence to substantiate, or refute, the claim that feedback enhances performance, and providing an estimate of the likely effect of these feedback interactions. No previous reviews had focused on the effects of feedback that result when a health professional educator and a health professional learner discuss the learner's clinical performance. A systematic review was used to search for and assess the evidence that this type of feedback might influence learner performance. High quality systematic reviews deliver the highest quality of evidence by assembling, critically appraising and synthesising all relevant randomised controlled trials, using a rigorous and systematic approach.[45] The following chapter (Chapter 2) describes this systematic review that specifically investigated the impact of verbal face-to-face feedback on health professionals' performance of a clinical task, compared to no or alternative feedback.

The systematic review was conducted according to Cochrane guidelines, involving multiple stages with exacting requirements. Designing and conducting the literature was challenging to start with. Developing inclusion and exclusion criteria was complex. The objective was to collect a cohesive set of studies that had been designed to investigate the effect of face-to-face feedback that is typical in clinical practice. Therefore clear criteria were needed to exclude studies with divergent targets such as an educator articulating feedback provided by a simulator, rather than feedback generated by the

educator based on their observation of the learner's performance; or assessment of team-based tasks or tasks that were not routinely part of clinical practice. In addition, when the literature search was initially conducted in November 2015, there were few eligible studies identified. It was then repeated in December 2017 and updated in February 2019 prior to publication. Each time, the search was conducted, more and higher quality studies were identified, indicating increasing interest and rigour in feedback research.

To succinctly summarise studies included in the review, a framework was developed to present extracted data on the interventions that had previously been shown to effect performance, such as instruction, practice and feedback components. This transformed a set of studies that reported information in disparate ways into an organised display that enabled clear and rapid comparison of key factors influencing study outcomes.

A number of software programs supported the processes in conducting this review, including 'EndNote' to manage literature search results; 'Covidence' to enable study selection by independent reviewers; and 'RevMan' to perform the meta-analysis and produce the required graphics.

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

This chapter is the study published in BMJ Open 2020;10:e030672. (Lightly edited for headings and formatting consistency). Published March 2020.

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

[1] Monash Doctors Education, Monash Health; Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria, Australia

[2] Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia

[3] Monash Health, Melbourne, Victoria, Australia

[4] Department of Physiotherapy, Monash University, Melbourne, Victoria, Australia

Key words: Feedback, Clinical practice, Delphi process, Health professions education, Educator behaviour

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

Funding: None

Declaration of interest: None

Acknowledgements: We would like to thank Professor Debra Nestel and Dr Kylie Horne for their comments on a draft of this work. We are indebted to the health professionals who generously videoed their feedback interactions.

2.2 Abstract

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

Design: Systematic review and meta-analysis.

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

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

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

Trial registration number:CRD42017081796

2.3 Background

Health professions education is embedded in clinical practice for both students and qualified staff as they continue learning and training.[22] Face-to-face verbal feedback focused on the performance of a clinical task involving an educator (senior clinician or peer) and a learner (any clinician) plays a crucial role in workplace learning, particularly within competency based education and programmatic assessment models.[46-49]

Multiple reviews on feedback in health professional education have been published, and include recommendations for effective practice.[50-53] Feedback can occur in various forms, including verbal, written or automated (for example, from a simulator or within an online learning module). The unique potential benefits of face-to-face verbal feedback are the opportunities for i) real-time interaction, to which the learner and educator bring their different perspectives, priorities and ideas to co-construct insights and strategies for improvement and ii) inter-personal connection, through which an educator can foster a learner's feelings of support, self-efficacy and motivation to improve, which are important catalysts in the learning process.[52,54-57]

There is widespread acceptance that feedback has an important role in maximising learning and achievement.[1,5,6,50] Ende[3] said, "Without feedback, mistakes go uncorrected, good performance is not reinforced, and clinical competence is achieved empirically or not at all." However there is little evidence to support this view that feedback enhances health professionals' performance. Indeed, a recent scoping review on feedback identified the need for systematic reviews to support evidence-based recommendations.[51]

The current strongest evidence relates to two systematic reviews that investigated the impact of audit and feedback. In 2006, Veloski *et al*[34] published a BEME systematic review in which almost 75% of included studies reported that audit and feedback could improve an individual physician's clinical performance, particularly when sustained and from an authoritative source. Feedback was defined as "summary information on clinical performance over a defined time period". They included any empirical study (not just

randomised controlled trials) and all types of physicians (most were primary care physicians). The majority of outcomes were clinical processes (such as test ordering) and the commonest data sources were medical records and billing records (none involved direct observation of performance).

In 2012, Ivers *et al*[33] updated a Cochrane review and meta-analysis that reported an increase in compliance with desired practice following audit and feedback, compared to usual care. The review included various health professionals (predominantly doctors), the unit of allocation for interventions ranged from individuals to health services, and the performance outcomes reported were clinical practice processes, such as the number or quality of prescriptions or tests. The authors argued that although the median risk difference (RD) in favour of feedback was small at 4.3% (interquartile range 0.5 to 16%), the 3rd quartile at 16% indicated that audit and feedback interventions could be much more effective. Using multivariable meta-regression, they identified that the effectiveness of audit and feedback increased when the source was a senior colleague or supervisor (RD 11%), the format involved both written and verbal components (RD 8%), the frequency was at least monthly (RD 7%), the aim was to reduce specific behaviour (RD 6%) and it included both explicit measurable targets and a specific action plan involving advice on how to improve, compared to performance information alone (RD 5%). In addition, two other factors were associated with a higher likelihood of improvement: a lower baseline performance and the type of behaviour being targeted, for example, prescribing (possibly perceived as 'important' and 'straightforward') had better outcomes than improving diabetes management (more complex) or test ordering (possibly perceived as 'less important').

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

Our research question was therefore:

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

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

2.4 Methods

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) statement.[58] The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) Registration ID CRD42017081796.

Eligibility criteria for considering studies for this review

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

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

All studies had to include at least one intervention involving verbal face-to-face feedback generated by a health professional, based on the observable performance of a workplace task performed by another health professional. A broad definition of feedback was permitted with a minimum requirement that it included information regarding learner performance. Studies were excluded if feedback was pre-determined or provided only by a simulated patient or machine. Audit and feedback studies, where feedback was based on aggregated quality performance indicators (such as numbers of tests ordered or degree of compliance with quality practice standards) were excluded,

as this was deemed to be distinctly different from a workplace task, such as suturing, that could be observed, objectively assessed and targeted for improvement with feedback. Two comparisons were evaluated 1) verbal face-to-face feedback compared with no feedback and 2) verbal face-to-face feedback compared to alternative feedback.

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

Search methods for identification of studies

We developed the search strategy in collaboration with a senior medical librarian using MEDLINE subject headings. Key words were used, including synonyms, truncation, wildcard and proximity operators related to 'feedback' AND 'health professional' AND 'performance' AND 'randomised controlled trial' (for the full search strategy for Ovid MEDLINE see supplementary information: S2.1). We translated this search strategy for other databases. The full holdings of Ovid MEDLINE (1946 to present with daily update), CENTRAL, Embase (1946 to present with daily update), CINAHL plus (1937 to present) and PsycINFO (1806 to present) were searched until 1st February 2019. We also searched the reference lists of systematic reviews and included studies.

Selection of studies

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

Data extraction

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

Assessment of risk of bias in included studies

The risk of bias was independently assessed by two authors (CEJ and JLK) for the selected performance outcome for individual studies, using Cochrane's 'risk of bias' tool (Chapter 8, Cochrane Handbook for Systematic Reviews of Interventions).[59] In particular, we used the following decision rules in assessing the risk of bias for specific individual domains. For 'participant and research team blinding': a participant receiving feedback or an educator giving feedback was deemed not to be blinded, even if they were deliberately not informed about the intervention or any differences between interventions. Nevertheless a 'low risk' rating was given if the outcome was not likely to be influenced by this lack of blinding, for example, if there were no changes to protocol or adherence that arose as a consequence of participant knowledge of group allocation.[60] For 'incomplete outcome data': to be rated as 'low risk', studies were required to include outcome data on at least 85% of the participants enrolled in each group (as per PEDRO guidelines),[61] and to provide participant numbers at the start and the number that dropped out during the study, from which group and the reasons.

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

Measures of treatment effect

Outcomes from included studies were expressed using point estimates and measures of variability (for example means (standard deviations SD) or medians (interquartile range IQR)). The effect was quantified using the standardised mean difference to combine

studies measuring the same outcome (task performance) using different measurement scales. When not reported, we estimated required data using available data or contacted study authors. If multiple outcomes were reported, we preferentially used the outcome that summarised multiple relevant task components, thereby providing a global, task-specific evaluation. If more than one reported outcome met this principle, we combined outcomes to provide a single metric using weighted averages of standardised scores.

We created and visually examined a funnel plot to explore reporting bias (Chapter 10, Cochrane Handbook).[62,63]

Data synthesis and assessment of heterogeneity

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

As a sensitivity analysis, we conducted a meta-analysis excluding studies with a high risk of bias. Using this pooled data, we calculated a prediction interval, which describes the range of likely results for new individual studies.[64]

We rated the overall certainty of evidence for the outcome using the GRADE approach (Chapter 12, Cochrane Handbook[65,66] and GRADE guidelines),[67] which considers within-study risk of bias, directness of evidence, heterogeneity, precision of effect estimates and risk of publication bias. Two authors independently rated the certainty of the evidence and resolved disagreements by discussion. We presented a summary of the evidence in a 'Summary of Findings' table.

Patient and Public Involvement

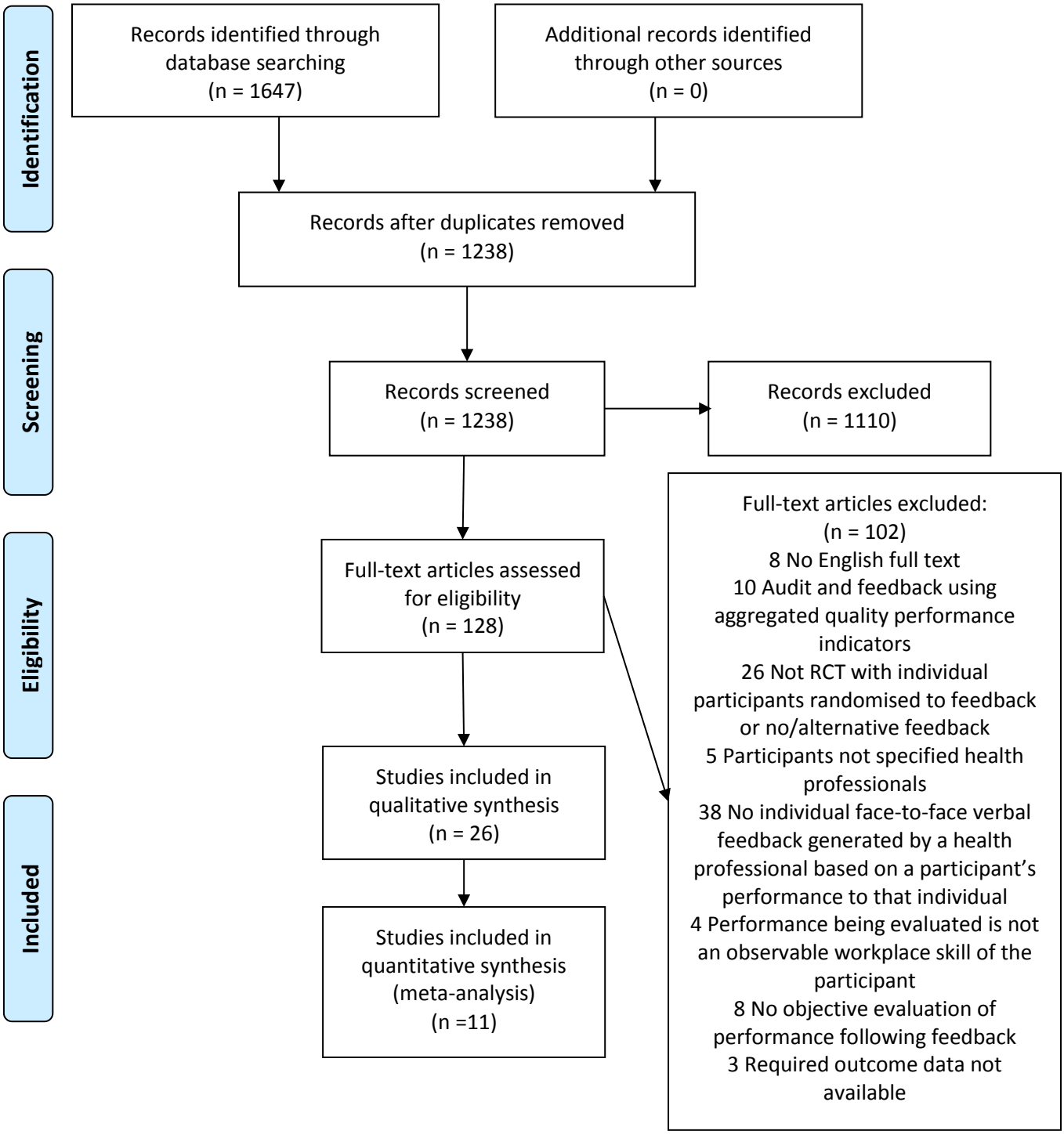
There was no involvement of patients or the public in any part of this research.

2.5 Results

Search results

The search yielded 1238 articles after 409 duplicates were removed. Based on title or abstract, we excluded 1110 articles. We assessed the remaining 128 full text articles for eligibility and found 26 randomised controlled trials that met all inclusion criteria. See Figure 2.1 for PRISMA study flow diagram.

Figure 2.1: PRISMA flow diagram for systematic review of verbal face-to-face feedback compared to no or alternative feedback.



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

Included studies

Table 2.1 presents the characteristics of included studies in this comparison. Eleven randomised controlled trials investigated the effect of verbal face-to-face feedback compared to no feedback on the objective evaluation of a workplace task. Seven (64%) reports were published in the last five years since 2014. The studies were conducted in Europe (4),[68-71] Canada (4),[72-75] the USA (2)[76,77] and Asia (1).[78]

There were 488 participants, including 196/366 (53.6%) men from seven studies that reported gender data.[68,70-73,76,78] Participants included 290 (60%) medical students in four studies,[68,71,75,76] 60 (12%) dental students in one study[70] and 138 (28%) doctors in six studies.[69,72-74,77,78] The workplace tasks involved a discrete task such as surgical procedures, cardiopulmonary resuscitation (CPR) or teaching skills, which occurred in clinical practice in four studies[72,73,77,78] and a simulation environment in seven studies (7/11, 64%).[68-71,74-76] Differences in feedback interventions between included studies involved feedback source (expert or peer), timing (during task performance, directly afterwards or delayed), content (evaluative information only or additional corrective advice, performance video, simulator information or written report) and number of feedback episodes. In addition, there was variation between studies in provision of instruction and expert demonstration of the task, opportunities for practice and duration of feedback intervention. (For more information, see supplementary information: S2.2 'Included studies' section).

Table 2.1: Summary of available data on characteristics of trials included in the comparison of verbal face-to-face feedback (intervention) compared to no feedback (control: no feedback from any external source) on performance

Author Year Country	**Task**	**Participants Health Profession Experience % Male**	**Teaching and Practice Same for Feedback Intervention and Control groups**	**Feedback Intervention**											**Study outcomes[a]** **All effects are SMD (95% CI). P value** in favour of feedback
				Additional information	Source		Timing			Content					
					Subject Expert	Peer	During task	Directly after	Delayed after	Verbal performance info	Verbal corrective advice	Machine output info[b]	Performance video	Written performance info	
Ahlborg 2015[68] Sweden	Simulated laparoscopic O&G surgery using a VR simulator (salpingectomy)	**Medical students** UGY5 50% M	Intervention duration: 1 session Case discussion + expert demonstration. 2 x practice trials. Performance evaluation: end of session.	2 x fb episodes. Fb given by expert i) during the task: fb given 'continuously, individualised by reinforcing & correcting each step' plus ii) directly after the task: fb based on simulator output information.	✓		✓	✓		✓	✓	✓			0.91 (-0.14 to 1.95) P = 0.08
Bonrath 2015[72] Canada	GI surgery in routine clinical practice (jejuno-jejunostomy during laparoscopic bariatric surgery)	**Doctors training in surgery** PGY3-5 72% M	Intervention duration: 2 months minimum. No teaching or practice in addition to routine clinical training. Performance evaluation: end of clinical attachment.	4 (approx.) x 25 min fb episodes. Fb given by expert using specific coaching model + video review of learner operating + video exemplars of good/poor technique. Effectiveness of strategies reviewed at subsequent session.	✓				✓	✓	✓		✓		1.62 (0.52 to 2.72) P = 0.002
Boyle 2011[69]	Simulated endovascula	**Doctors training in surgery**	Intervention duration: 1 session	5 x fb episodes. *Experts* provided 'whatever	✓			✓		?	?	✓			1.27 (-0.32 to 2.87)

(expert fb) Ireland	r surgery using a VR simulator (renal artery angioplasty + stent)	PGY4+	Teaching + expert demonstration. 5 x practice trials. Performance evaluation: end of session	feedback they considered appropriate' + simulator output information.											P = 0.08
Boyle 2011[69] (peer fb)	Same as above	Same as above	Same as above	5 x fb episodes. *Peer* discussed simulator output, any task errors & teaching instructions given at start.		✓		✓		✓	✓	✓			0.81 (-0.66 to 2.29) P = 0.24
Kroft 2017[73] Canada	O&G surgery in routine clinical practice (laparoscopic salpingectomy)	**Doctors training in O&G** PGY2-6 33% M	Intervention duration: 1 x 15min practice using laparoscopic salpingectomy module on VR surgical simulator within 1h of surgery. Performance evaluation: laparoscopic salpingectomy in OR soon afterwards.	1 x fb episode from expert directly after VR simulator practice. Fb 'standardized and given in an evidence based fashion to optimise effectiveness' & included '3 constructive recommendations based on performance'.	✓			✓		✓	✓				0.85 (-0.35 to 2.06) P = 0.14
O'Connor 2008[76] USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	**Medical students** UGY1-2 44% M	Intervention duration: 4 wk. 2h instruction + practice suturing & knot tying until able to do it easily. Then instruction on laparoscopic surgery + expert demonstration video of task tying, followed by 30 mins familiarisation with equipment.	Expert fb provided 'continually on how to improve' during practice sessions + detailed explanations of simulator output information at the end of the session + given target performance goals.	✓		✓	✓		✓	✓	✓			0.40 (-1.25 to 2.04) P = 0.58

			Practice: 1h daily, 6 days per week for 4 weeks. Performance evaluation: combined assessment of each attempt throughout intervention.												
Olms 2016[70] Germany	Simulated colour matching teeth	**Dental students** UGY3	Intervention duration: 1 session Study conducted during 10 wk routine university module on matching tooth shades involving variety of teaching + practice opportunities. Performance evaluation: 2 wks after intervention (within one university module).	1 x expert fb session. Fb included correct response + explanation with expert demonstration if needed + written copy of evaluation. Expert trained in fb.	✓			✓		✓	✓			✓	2.09 (1.45 to 2.73) P < 0.001
Pavo 2016[71] Austria	Simulated CPR	**Medical students** UGY3 57% M	Intervention duration: 1 session. Instruction on basic life support occurred previously, as part of university course. 1 x 2h additional training session: instructional video + training using modified Peyton 4 step approach.[c] Brief practice (few mins) in pairs using a manikin. Performance evaluation: end of session.	Fb during performance from peer performing ventilation to the student performing compressions (being assessed), at the start of each set of 30 chest compressions. Fb included information + corrective advice on compression rate & depth, hand position, decompression & hands-off time. Instructional video for intervention group had demonstrated this.		✓	✓			✓	✓				0.25 (-0.02 to 0.51) P = 0.06

Skeff 1983[77] USA	Clinical teaching skills during ward round	**Physicians**	Intervention duration:1 session in the middle of 4wk ward duty. At mid & end of ward duty: video of physician's teaching on ward rounds + rating of physician's teaching skills by medical students and junior medical staff on ward (video + ratings not shown to control group). Performance evaluation: 2 wk later, at end of ward duty	1 x 60 min fb discussion with peer, including video review, trainee ratings & self-assessment to enable physician to identify strengths & devise solutions to problems.	✓				✓	✓	✓		✓	✓	0.56 (-0.15 to 1.27) P = 0.12
Soucisse 2017[74] Canada	Simulated surgical procedure (bench-top intestinal anastomosis using cadaveric dog bowel)	**Doctors training in surgery** PGY1-4	Intervention duration: 1 session. Task instruction occurred previously (no teaching or practice within intervention). Baseline performance videoed. Performance evaluation:3 wk later (ongoing clinical work as a surgical resident).	1 x 30min expert fb sometime after baseline performance with video review of baseline performance + adapted coaching model including 2-3 suggestions for improvement + expert demonstration followed by learner demonstration of desired improvements, as required + action plan.	✓				✓	✓	✓		✓		0.3 (-0.44 to 1.05) P = 0.42
Vafaei 2017[78] Iran	Chest ultrasound for trauma patients in Emergency	**Doctors training in emergency** PGY4 57% M	Intervention duration: 1 session. Instruction for task occurred in previous training year (no teaching or practice within intervention).	1 x 5min expert fb, directly after baseline performance assessment, on 'weak and strong points' and based on specific procedural skill assessment checklist.	✓			✓		✓					3.04 (1.95 to 4.13) P < 0.001

			Baseline performance assessed. Performance evaluation: 2 months later (ongoing work as emergency resident).												
Xeroulis 2007[75] (fb after) Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	**Medical students** UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. Performance evaluation: end of session.	Expert fb as needed (expert or learner initiated), after practice trials, involving constructive ways to improve + expert demonstration.	✓			✓		✓	✓				0.86 (-0.08 to 1.80) P = 0.06
Xeroulis 2007[75] (fb during)	Same as above	Same as above	Same as above	Same as above except expert fb during practice trials.	✓		✓			✓	✓				1.44 (0.43 to 2.46) P = 0.004

Footnotes:

a= See 'Meta-analysis' section in Results for additional study details.

b= Machine output information: simulator metrics (e.g. procedural time or instrument path length) or CPR machine information (e.g. compression rate and depth)

c= Peytons' 4 step model[79]

Abbreviations:

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

Risk of bias

The risk of bias graph is presented in Figure 2.2 and the risk of bias summary is presented in Figure 2.3. In summarising the risk of bias across domains within each study, two studies were rated 'low risk',[72,74] six studies were rated 'unclear'[68,69,71,73,75,77] and three studies were 'high risk'.[70,76,78] (For more information, see supplementary information: S2.2 'Risk of bias' section).

Figure 2.2: Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies

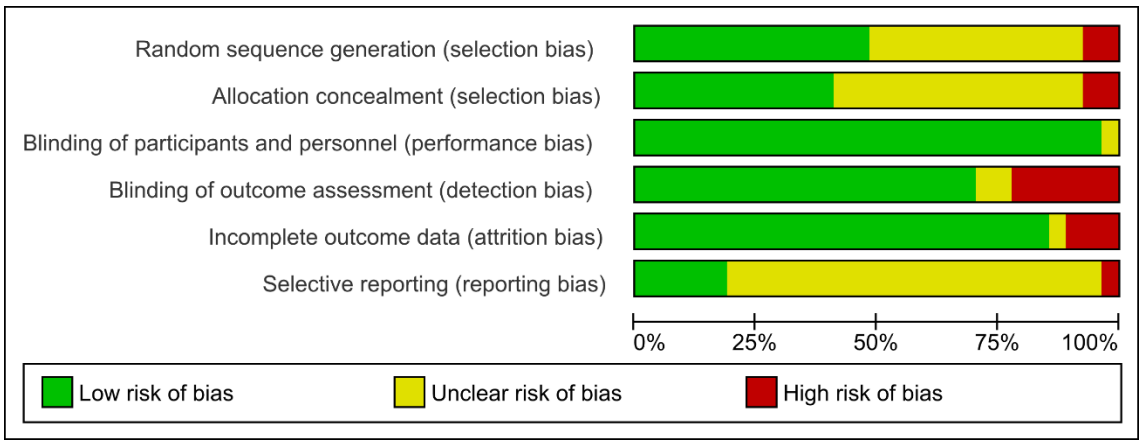
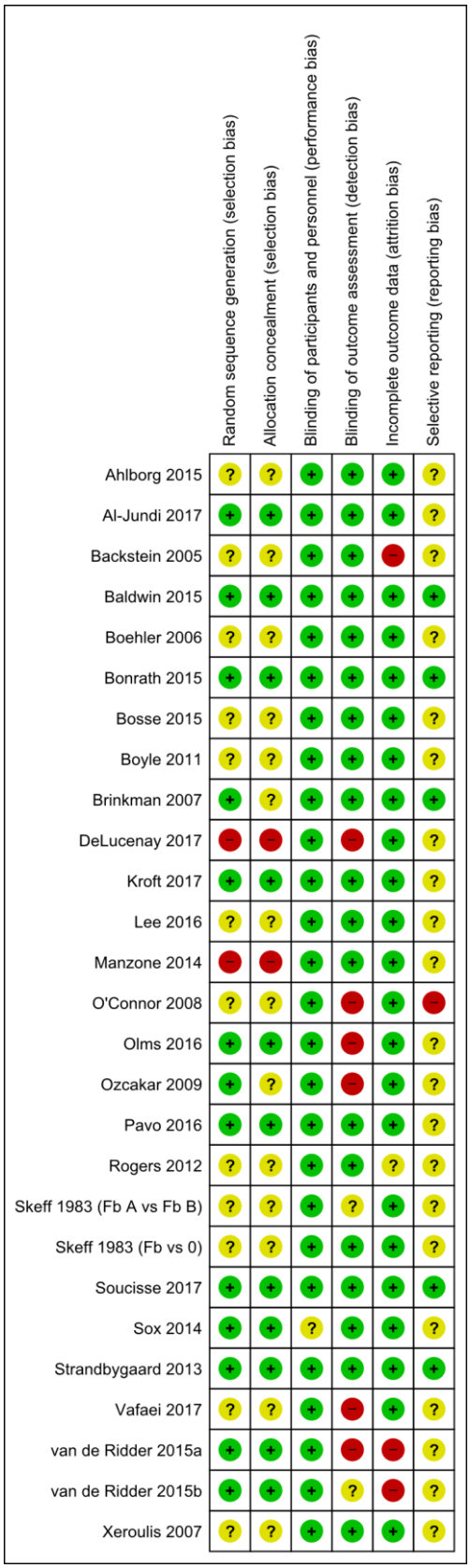


Figure 2.3: Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

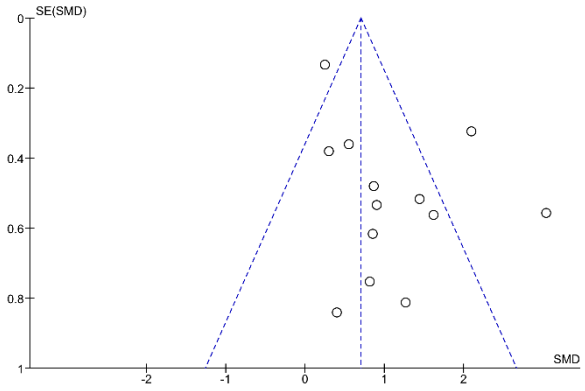


Reporting bias

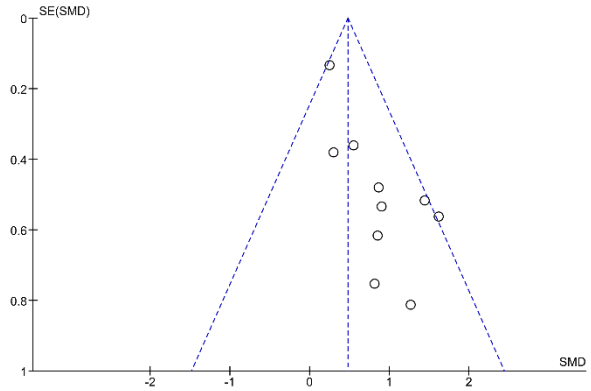
The funnel plots are presented in Figure 2.4: for all included studies (Figure 2.4a) and after excluding studies at high risk of bias (Figure 2.4b). Both funnel plots are asymmetrical, with a paucity of small studies with negative effect sizes that are less likely to be published, indicating some potential for publication bias.

Figure 2.4: Funnel plot of the comparison of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

a) all included studies



b) excluding studies at high risk of bias



Abbreviations:

SE = standard error; SMD = standardised mean difference

Footnote:

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

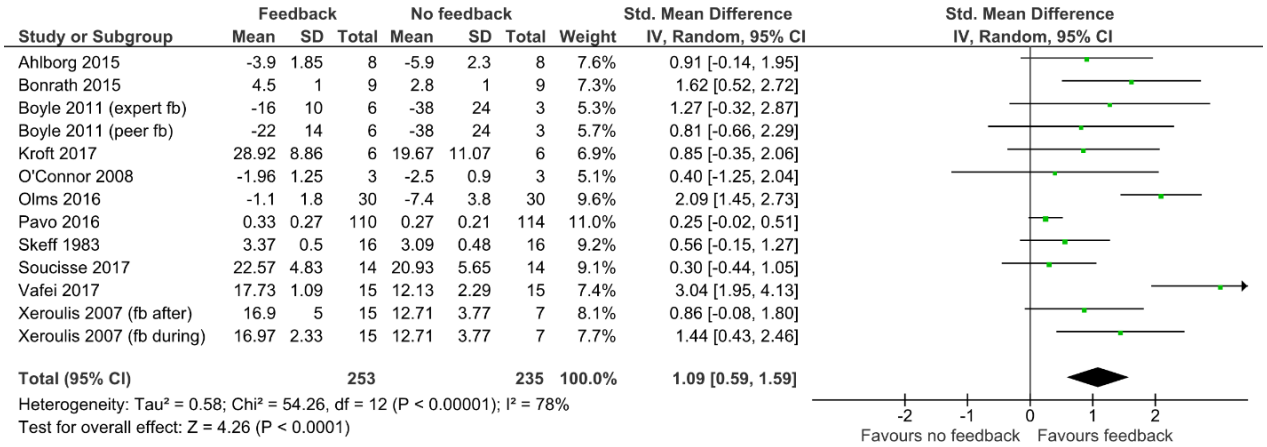
Meta-analysis

A meta-analysis of the impact of verbal face-to-face feedback compared with no feedback on performance included 13 comparisons from the 11 studies, involving 488 participants. Two studies reported data that each enabled two comparisons: in one study, feedback provided *during* practice in one group and *directly after* practice in another were compared with no feedback;[75] in another study, feedback provided by an *expert* in one group and by a *peer* in another group[69] were compared with no feedback. In the meta-analysis, numbers for the control group for these studies were halved to retain sample independence.[65]

The meta-analysis of the effect of verbal face-to-face feedback compared with no feedback on workplace task performance found a standardised mean difference of 1.09 (95% confidence interval (CI) 0.59 to 1.59; P < 0. 001) using a random effects model. The forest plot is presented in Figure 2.5a.

Figure 2.5: Forest plot for the meta-analysis of the effect of verbal face-to-face feedback, compared to no feedback, on performance.

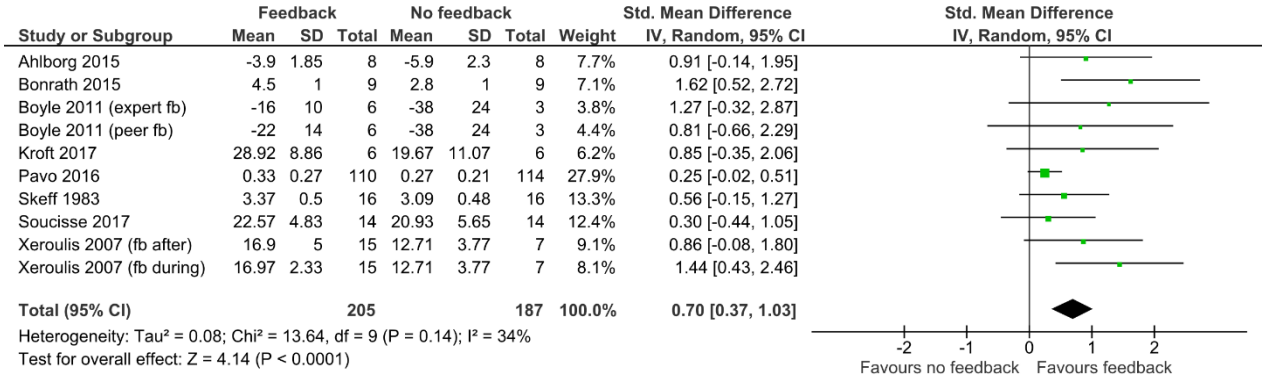
a) All included studies



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

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

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



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

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

Abbreviations:

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

Footnotes:

Ahlborg 2015: mean and SD read from graph
Boyle 2011: mean and SD read from graph
Bonrath 2015: combined outcome calculated
Pavo 2016: median taken as best estimate of mean and calculated SD from IQR
Xeroulis 2007: SD estimated from 95% CI

Sensitivity analysis

As a sensitivity analysis, we repeated the random effects meta-analysis after excluding studies with a high risk of bias. Eight studies (8/11, 73%) were included that involved 392 health professional learners across ten comparisons.[68,69,71-75,77] The standardised mean difference was 0.7 (95% CI 0.37 to 1.03; P < 0.001). The forest plot is presented in Figure 2.5b. The prediction interval was -0.06 to 1.46.

We judged that the certainty of the evidence was low, using the GRADE approach. We downgraded the overall rating from high to low, in view of a serious risk of bias (in particular, due to a lack of concealment and potential for selective reporting of outcomes) and publication bias.[80] For more information, see supplementary information: S2.2 'Certainty of evidence' section). Figure 2.6 displays the Summary of Findings table.

Figure 2.6: Summary of findings table for the effect of verbal face-to-face feedback, compared to no feedback, on performance.

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

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

Included studies

Table 2.2 presents the characteristics of included studies in the comparison of verbal face-to-face feedback compared to alternative feedback. Twenty studies (22 comparisons) were included in this analysis and involved verbal face-to-face feedback compared to alternative feedback. Nine studies (9/20, 45%) were published in the last five years since 2014. The studies were conducted in Europe (8/20, 40%), USA (7/20, 35%), Canada (4/20, 20%), and Asia (1/20, 5%).

There were 1974 participants, including 660/1463 (45%) men from 13 studies that reported gender data.[71,76,81-91] Included studies involved students (medical, mixed health professions and pharmacy) (1869, 95%) in 16 studies,[71,75,76,81-84,86-94] and doctors (105, 5%) in four studies.[69,77,85,95] All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[77,85] Three studies evaluated performance in a clinical practice setting (involving teaching skills,[77] professional and communication skills[85] and oral case presentations)[94] and the remaining 17 assessed performance in a simulated environment (surgical procedures, nasogastric tube insertion, intubation, hearing test, pharmacy consultation or CPR).[69,71,75,76,81-84,86-93,95] (For more information, see supplementary information: S2.3 'Included studies' section).

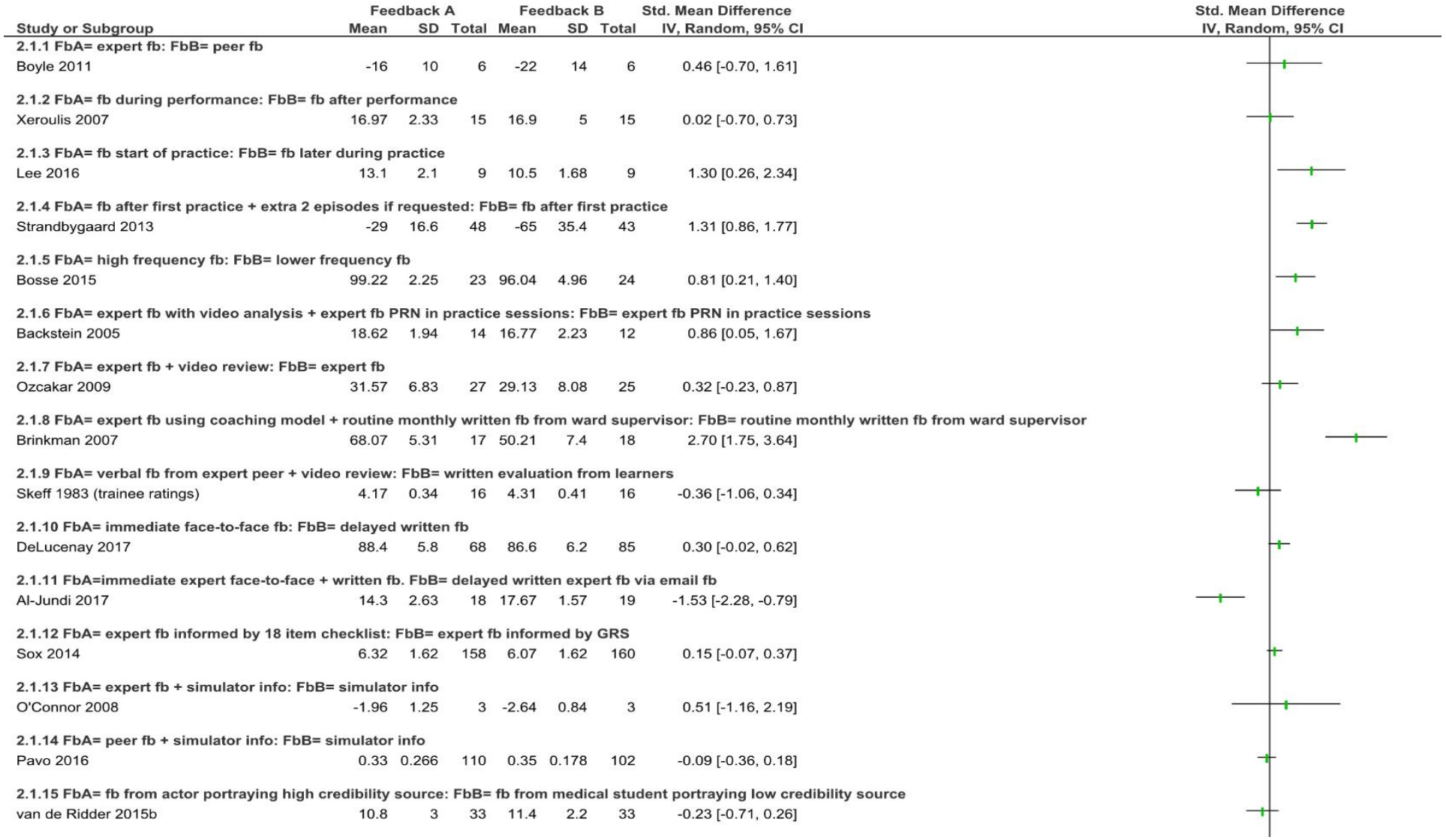
Risk of bias

In summarising the risk of bias across domains within each study, two studies were rated as low risk,[82,89] seven studies were rated as 'high risk',[76,86,90,91,93,95] and the remaining studies were rated as 'unclear'. (See the risk of bias summary in Figure 2.3). (For more information, see supplementary information: S2.3 'Risk of bias' section).

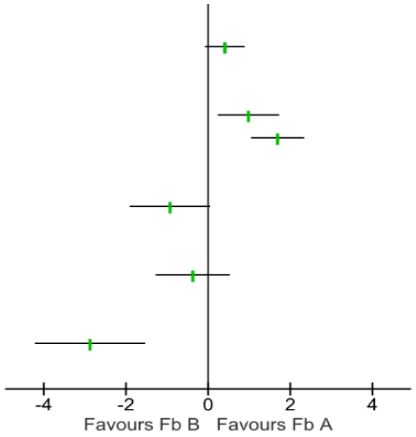
Effect of interventions

Figure 2.7 presents the forest plot and standardised mean differences (SMD). One additional study[82] that reported categorical data is not included in the forest plot. It compared a learning conversation (315 participants, pass rate 80.9%) to a feedback sandwich (325 participants, pass rate 77.2%) resulting in an odds ratio of 1.25 (95% CI 0.85 to 1.84) that favoured the learning conversation. The feedback comparisons were markedly diverse, so we did not pool outcomes in meta-analysis.

Figure 2.7: Forest plot for the effect of verbal face-to-face feedback (Feedback A), compared to alternative feedback (Feedback B), on performance.



2.1.16 FbA= corrective fb using positive language: FbB= corrective fb using critical langauage							
van de Ridder 2015a	7.5	1.3	37	6.9	1.6	36	0.41 [-0.06, 0.87]
2.1.17 FbA= expert fb: FbB= compliment							
Boehler 2006	21.98	3.48	16	17	6.01	17	0.98 [0.25, 1.71]
Rogers 2012	8.4	2.2	27	5.1	1.6	27	1.69 [1.06, 2.32]
2.1.18 FbA= verbal info, focused on task: FbB= verbal info, compared to others							
Manzone 2014 (verbal/task vs verbal/others)	2.89	1	10	3.86	1	9	-0.93 [-1.89, 0.03]
2.1.19 FbA= verbal info, focused on task: FbB= numerical info, focused on task							
Manzone 2014 (verbal/task vs number/task)	2.89	1	10	3.28	1	10	-0.37 [-1.26, 0.51]
2.1.20 FbA= verbal info, focused on task: FbB= numerical info, compared to others							
Manzone 2014 (verbal/task vs number/others)	2.89	1	10	5.89	1	10	-2.87 [-4.20, -1.55]



Abbreviations:

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

Footnotes:

[1]Baldwin 2015: categorical data not included in this figure; see text in Results.
Al-Jundi 2017: additional information (data to calculate mean and SD for each group) from author.
Boehler 2006: additional information (number of participants in each group and SD) from author.
Lee 2016: calculated SD from SE.
Manzone 2014: calculated standardised score to combine outcome of supine and normal positions.
Pavo 2016: median taken as best estimate of mean.
Rogers 2012: additional information (standard deviation) from author.
Sox 2014: SD derived from reported t, p and mean values. Assumption that SDs were equivalent for intervention and controls.
Strandbygaard 2013: SE derived from 95% CI

Table 2.2: Summary of available data on characteristics of trials comparing the effect of verbal face-to-face feedback (Intervention A), to alternative feedback (Intervention B), on performance.

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
First author Year Country		Health profession Experience Gender: % Men		All included verbal face-to-face feedback to an individual health professional		Unless otherwise stated, effects are SMD (95% CI). P value favours feedback intervention A
Al-Jundi[b] 2017[81] England	Simulated surgical skill using bench top model ('skin' suturing with a latex pad)	**Medical Students** UGY5 65% M	Intervention duration: 1 session Video instruction on surgical skill. 1 x 10 mins for baseline performance. Performance evaluation: 2 days later	**Immediate face-to-face + written expert feedback** 1 x expert fb. Expert observed baseline performance and rated it using task-specific checklist. Learner completed written self-assessment using same check list. Fb directly after performance, by expert with medical education qualification. Fb included verbal fb based on assessment checklist, 'directive and specific' + demonstration of skill, as required. Learner given copy of assessment + written feedback forms.	**Delayed written expert fb via email** 1 x written expert fb via email same day as baseline performance. Expert watched video of baseline performance, rated it using task-specific checklist and wrote fb comments aligned with assessment checklist, including suggestions for improvement, so fb was 'directive and specific'. Both assessment and written feedback forms emailed to learner.	-1.53 (-2.28 to -0.79) P < 0.001 favours feedback intervention B
Backstein 2005[95] Canada	Simulated surgical procedure using a bench top model (vascular anastomosis)	**Doctors in surgical training** PGY1	Intervention duration:4 wk Lecture on surgical procedure. 3 x 2h weekly practice sessions with expert fb as needed. Expert vascular surgeons undertook fb	**Review of performance video with expert fb** **+ practice sessions with expert fb available** 3 x weekly videotaping of surgical procedure, with	**Practice sessions with expert fb available**	0.86 (0.05 to 1.67) P = 0.03

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
			training, based on evaluation checklist and given in a similar manner. Performance evaluation: in wk 4	expert feedback available during task, followed by up to 15min review of video with expert fb.		
Baldwin 2015[82] England	Simulated BLS	**Health professional students** medical (58%), physio (12%), pharmacy (10%), nursing (10%), dentistry (10%). UGY1 33%	Intervention duration: 4 wk Instruction and practice with manikin 3 x 2.5h weekly. Fb provided directly after performance by senior peer instructor. Instructor accredited in BLS + trained to provide fb. Fb provider compliance monitored. Performance evaluation: in wk 4	**'Learning conversation' model** Fb focused on learner's perspective: started with learner self-assessment, then explored issues and ideas raised by learner with group using advocacy inquiry format[c] with final summary.	**'Feedback sandwich model'** Fb involved a point for improvement in between 2 points of praise.	OR 1.25[d] (0.85 to 1.84) P = 0.25
Boehler 2006[83] USA	Simulated surgical skill using a bench top model (tying a 2-handed square knot)	**Medical students** UGY2-3 52% M	Intervention duration: 1 session Instruction in knot tying from surgeon. 1 x baseline performance. Performance evaluation: end of session.	**Expert feedback** 1 x episode of fb from expert surgeon, directly after performance, describing 1-2 specific ways to improve performance.	**Compliment** 1 x pre-scripted general compliment e.g. 'great job!'	0.98 (0.25 to 1.71) P = 0.01
Bosse 2015[84] Germany	Simulated nasogastric tube insertion (NGTI) into manikin	**Medical students** UGY1-2	Intervention duration: 1 session	**High frequency fb** 6 x episodes of fb, given after each practice trial.	**Low frequency practice** 2 x episodes of fb, given after first and last practice trial.	0.81 (0.21 to 1.40) P = 0.01

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
		51% M	NGTI training using case study role play and 4 step procedural training method[e] 6 x practice trials. Fb 'positively worded', focused on effect of actions, given directly after performance by senior peer instructors, trained in procedure & fb. Performance evaluation: end of session.			
Boyle 2011[69] Ireland	Simulated endovascular surgical procedure using a VR simulator (renal artery angioplasty + stent)	**Doctors training in surgery** PGY4+	Intervention duration: 1 session Teaching & expert demonstration. Fb providers had simulator training. 5 x practice trials (each maximum 40min). Performance evaluation: end of session.	**Expert fb** 5 x fb episodes. Experts provided 'whatever feedback they considered appropriate' and simulator output information.	**Peer fb** 5 x fb episodes. Peer discussed simulator output, any task errors & the teaching instructions given at start of session.	0.46 (-0.70 to 1.61) P = 0.41
Brinkman 2007[85] USA	Professional and communication skills during routine clinical practice on a paediatric ward	**Doctors training in paediatrics** PGY1 34% M	Intervention duration: 1 session No teaching or practice within intervention Routine feedback as part of clinical training: monthly written evaluations from paediatricians on ward duty.	**Coaching session + routine feedback as part of clinical training** 1 x 30min fb session soon after initial evaluation at start of attachment, based on summarised performance ratings from nurses & parents.	**Routine feedback as part of clinical training** Performance ratings from nurses and patients not seen.	2.70 (1.75 to 3.64) P < 0.001

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
			Performance ratings obtained from nurses and patients at start and end of doctors' rotation. Performance evaluation: 5 months after start of clinical attachment.	Used a coaching model focused on assisting learner to understand information, design goals and improvement strategies. Fb given by paediatricians trained in coaching model.		
DeLucenay 2017[86] USA	Simulated pharmacist patient consultation (identifying prescription errors and communication skills)	**Pharmacy students** UGY3	Intervention duration: 1 semester. Study conducted during usual university module on medication counselling involving 15 min SP consultations, each on a different topic. Directly after each one, SP provided 5min fb on communication skills. Performance evaluation: last 4 SP consultations.	**Immediate face-to-face fb** 4 x expert fb directly after SP consultation and SP fb, based on expert's direct observation of SP consultation (unseen by participants). Fb included performance grade, performance and topic discussion with suggested improvements.	**Delayed written fb** 4 x videotaping of SP consultation. Expert reviewed video then provided written fb and grade via intranet, prior to next practice.	0.30 (-0.02 to 0.62) P = 0.07
Lee 2016[87] Canada	Simulation urological surgical procedure using a bench top model (flexible ureteroscopy for urolithiasis)	**Medical students** UGY3-4 78% M	Intervention duration: 3 wk Instruction and expert demonstration of procedure, followed by 3 x weekly 30min practice sessions. Performance evaluation: end of 3rd session.	**Early feedback** 1 x 10-15min expert fb directly after first practice attempt, focused on assessment domains.	**Late feedback** Same as early fb but at end of second practice session.	1.3 (0.26 to 2.34) P = 0.01

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
Manzone 2014[93] (verbal comment focused on performance vs verbal comment + comparison to training levels) Canada	Simulated intubation using manikin	**Medical students** UGY1-2	Intervention duration: 1 session Instructional video on intubation. 1-1.5h practice with manikin, with learner in 4 different positions (5 x practice trials in each position). 10 x fb by expert, given directly after practice trials in 2 positions (2 x 5). Fb only provided performance evaluation, with no advice on how to improve. Performance evaluation: end of session.	**Performance comment focused on task** Fb involved evaluative performance comment, focused on any 2 aspects of performance (either done correctly or not) e.g. 'improper use of the laryngoscope'. + individual's progress on task.	**Performance comment compared to others (different training levels)** Fb involved evaluative performance comment, focused on comparison of learner's performance with expected standards at different training levels e.g. 'your performance was at the level of a resident.'	-0.93 (-1.89 to 0.03) P = 0.05 favours fb intervention B
Manzone 2014[93] (verbal comment on performance vs numerical rating, focused on individual progress)	As above	As above	As above	**Performance comment focused on task** As above	**Numerical performance outcome, focused on task progress** Provided with numerical performance information (performance time and number of hand movements). Plotted on graph to focus on own progress.	-0.37 (-1.26 to 0.51) P = 0.39 favours feedback intervention B
Manzone 2014[93]	As above	As above	As above	**Performance comment focused on task**	**Numerical performance outcome, compared to**	-2.87 (-4.20 to -1.55)

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
(verbal comment focused on performance vs numerical fb + comparison)				As above	**others (scores at different training levels)** Provided with numerical performance information (performance time and number of hand movements), accompanied by a list of scores across different training levels from medial student to specialist.	P < 0.001 favours feedback intervention B
O'Connor 2008[76] USA	Simulated surgical skill using a laparoscopic simulator (suturing & knot tying)	**Medical students** UGY1-2 44% M	Intervention duration: 4 wk 1st session: 2h instruction and practice suturing & tying knots 'until able to do it easily'. 2nd session: instruction on laparoscopic surgery and expert demonstration video on task, followed by 30mins familiarisation with equipment. Practice: 1h daily, 6 days per week for 4 weeks Simulator output information available at the end of each practice session: task completion time, smoothness of tool	**Expert fb during practice + simulator output information with expert discussion** Fb by surgical expert occurred continually throughout practice sessions. Expert observed participants closely, corrected mistakes early and provided instructions on how to improve. + simulator output information with expert explanation of this information & given target goals.	**Simulator output information**	0.51 (-1.16 to 2.19) P = 0.48

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
			manipulation and path length of tool.			
Ozcakar 2009[88] Turkey	Simulated patient consultation with a simulated patient (communication and history taking skills)	**Medical students** UGY2 62% M	Intervention duration: 2 wk Study conducted during routine university module on clinical skills training. Evaluation: 2 wk after intervention following clinical skills lectures + practice with video recording.	**Video review with expert + expert fb** 1 x videotaping of SP consultation. Directly afterwards, review video with expert plus fb.	Expert fb 1 x expert fb directly after SP consultation	0.32 (-0.23 to 0.87) P = 0.24
Pavo 2016[71] Austria	Simulated CPR	**Medical students** UGY3 57% M	Intervention duration: 1 session Instruction on basic life support occurred previously, as part of university course. 1 x 2h additional session including training using modified Peyton 4 step approach[e] and practice on a manikin. Performance evaluation: CPR skills at end of session.	**Verbal fb from peer during CPR** Fb during performance from peer performing ventilation to the student performing compressions (being assessed), at the start of each set of 30 chest compressions. Fb included information + corrective advice on compression rate & depth, hand position, decompression & hands-off time. Brief practice by pair of participants with a manikin, until felt confident.	**Machine output during CPR** CPR machine showed real time visual display (numbers and graphs) of compression rate & depth plus automated audio advice to correct any deviations during CPR.	-0.09 (-0.36 to 0.18) P = 0.53 favours feedback intervention B

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
Rogers 2012[92] USA	Simulated surgical skill (tying a single 2-handed square knot)	**Medical students** 'surgical clerkship year'	Intervention duration: 1 session Training in knot tying. 2 x practice trials (1 before & 1 after training). Performance evaluation: end of session.	**Expert fb** 1 x fb from expert, with specific information on improving subsequent performance, directly after performance.	**Compliment** 1 x general compliment from expert, instead of fb.	1.69 (1.06 to 2.32) P < 0.001
Skeff 1983[77] USA	Clinical teaching skills during ward round in routine clinical practice	**Attending Physicians**	Intervention duration: 1 month Performance evaluation: medical students and junior medical staff (trainees) on ward rated physicians' teaching skills during ward rounds, at the mid- and end of 1 month term.	**Expert peer fb** 1 x 1h session mid-term with expert peer, including review of videos of physician's teaching on ward rounds, trainees' evaluations and self-assessment of teaching skills. Fb discussion aimed at helping physician clarify strong teaching skills and devise solutions for teaching problems	**Written fb** Received written summary of trainees' evaluation of physician's teaching skills.	-0.36 (-1.06 to 0.34) P = 0.30 favours feedback intervention B
Sox 2014[94] USA	Case presentation during student clinical attachment in paediatrics	**Medical students** UGY3	Intervention duration: paediatric clerkship Week 1: Lecture on important aspects of case presentations. Week 2: present case to small group with doctor in paediatric unit who was trained in evaluation.	**Detailed evaluation form** 1 x constructive expert fb, directly after performance informed by 18 item evaluation form. Learner saw 18 item evaluation form but not given a copy.	**Simple evaluation form** 1 x constructive expert fb, directly after performance informed by single item GRS evaluation form. Learner saw 1 item evaluation form but not given a copy.	0.15 (-0.07 to 0.37) P = 0.17

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
			Performance evaluation: end of clerkship			
Strandbygaard 2013[89] Denmark	Simulated O&G surgery using a VR laparoscopic simulator (salpingectomy for extra-uterine pregnancy)	**Medical students** UGY 4-6 44% M	Intervention duration: 2 months 1 x session with instruction + expert demonstrations on operational technique, how to use simulator and interpret simulator output information. Simulator output information available after every practice: procedural time + performance score derived from multiple task performance criteria. Participants instructed to practice until achieved predefined expert proficiency level; could practice daily (max 3h) for up to 2 months.	**Standardised expert fb with later, additional expert fb if requested by learner + simulator performance score** 1-3 x 10-12min episodes of expert fb involving information on how to perform task components correctly. 1st fb episode provided after first practice trial; learner could ask for up to 2 additional fb episodes (optional) involving same standardised advice + simulator performance score.	**Simulator performance score**	1.31 (0.86 to 1.77) P < 0.001
Van de Ridder 2015a[90] (Advances in Health Science Education) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 35% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. Fb from senior medical student with acting experience & trained to act	**Positively framed fb** 1x fb directly after baseline performance. Fb comment started with global praise followed by the most suitable suggestion for improvement,	**Negatively framed fb** 1x episode fb directly after practice trial. Fb comment started with global criticism followed by most appropriate directive advice for improvement,	0.41 (-0.06 to 0.87) P = 0.08

Article	Task	Participants:	Common to interventions A + B	Intervention A	Intervention B	Study outcomes[a]
			as a physician familiar with W&R test. Fb provider trained to give corrective information, cast in positive or negative tone according to study group allocation. Performance evaluation: end of session.	selected from a list of 4 commonest task errors (e.g. 'You did this well; a tip is …')	selected from list of 4 commonest task errors. (e.g. 'You did not do this correctly; you should change.')	
Van de Ridder 2015b[91] (Medical Teacher) Netherlands	Simulated hearing test with a simulated patient (Weber & Rinne test)	**Medical students** UGY1 31% M	Intervention duration: 1 x session Instructional video of task. 1 x baseline performance. All fb providers trained for 1h on W&R test and giving fb according to protocol. Fb monitored to ensure it was given as per protocol. Performance evaluation: end of session.	**High credibility fb provider** 1 x fb directly after performance comprised of 2 points for improvement from actor portraying high credibility fb provider (operationalised as older, male, name tag & introduced as Professor ENT, wearing a white coat).	**Low credibility fb provider** 1 x fb directly after performance comprising 2 points for improvement from senior medical student portraying low credibility fb provider (operationalised as young, female, informally dressed).	-0.23 (-0.71 to 0.26) P = 0.36
Xeroulis 2007[75] Canada	Simulated surgical skill using a bench-top model (suturing & knot tying)	**Medical students** UGY1	Intervention duration:1 session Instructional video on task. Practice involved 19 x trials in 1h. Fb involved constructive ways to improve + expert demonstration. Performance evaluation: end of session.	**Expert fb during practice** Expert fb as needed (expert or learner initiated), *during* practice trials.	**Expert fb directly after practice** Same as 'during practice' except fb *after* practice trials.	0.02 (-0.70 to 0.73) P = 0.96 favours feedback intervention B

Abbreviations:

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

Footnotes:

a= See Figure 2.7 forest plot for additional study details;

b = additional data obtained from authors, enabling calculation of mean, SD and % men;

c = Advocacy Inquiry approach;[96]

d = Categorical data only available (see text in Results for more details);

e = Peyton's 4 steps.[79]

2.6 Discussion

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

Our meta-analysis found that verbal face-to-face feedback may result in a moderate to large improvement in health professionals' performance compared to no feedback, with SMD 0.7 (95% CI 0.37 to 1.03; P < 0.001) from eight studies involving 392 health professionals, after excluding studies at high risk of bias. However, the quality of evidence was low, primarily due to risk of bias and publication bias. To our knowledge, this is the first report to provide some substantiation for the widely held view that feedback enhances health professionals' performance and to estimate the benefit. (For more information, see supplementary information: S2.4 'Discussion: Comparison 1' section).

The consistent positive effects across all included studies, with substantially overlapping confidence intervals, supports the likelihood that verbal face-to-face feedback enhances performance in the health professions. Our pooled effect size was moderate to large at 0.7.[97] The calculated prediction interval for the comparison of verbal face-to-face feedback with no feedback (excluding studies with a high risk of bias) was -0.06 to 1.45. This indicates a wide range in the likely feedback effect for any individual study, from a very small detrimental impact to a very large beneficial effect, on performance. These results align with previous meta-analyses within health and other professions that have reported beneficial but variable effect sizes with different feedback interventions.[1,33,34] For example, within the health professions, Ivers *et al*[33] reported that 0.5 to 16% more participants followed desired practice when involved in an audit and feedback intervention. In comparison, a meta-analysis by Kluger and DeNisi,[2] which analysed any type of feedback across any discipline, compared with no feedback, reported a pooled SMD of 0.4; notably one third of included studies reported a detrimental impact.

One possible explanation for this variability, is that some constituents within a feedback intervention are more effective than others. When specific feedback elements were isolated, the largest beneficial effects of feedback reported in Kluger and DeNisi's[2] meta-

analysis were i) effect size 0.55 when feedback included information on any changes since the previous attempt, ii) effect size 0.51 when a specific and challenging goal was set, iii) effect size 0.47 when feedback posed little threat to self-esteem and iv) effect size 0.43 when feedback included information on the correct outcome.

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

For the second comparison of the effect of verbal face-to-face feedback compared to alternative feedback on performance, there was a diverse range in the alternative feedback interventions, which precluded meta-analysis. Where individual studies tested the relative impact of different feedback interventions, there was greater performance improvement seen with the following strategies: additional expert coaching sessions compared to routine monthly written feedback from supervisors;[85] expert feedback early in a practice period compared to later;[87] additional episodes of feedback from experts;[84,89] additional episodes of feedback involving expert video analysis[95] and expert feedback compared to compliments.[83,92]

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

The studies assembled in this review illustrate the wide variety of possible constituents within feedback interventions and the potential influence on performance. Within verbal face-to-face feedback interventions, there were important differences between included studies in feedback content, source and timing. Previous studies have noted potential beneficial effects attributable to feedback that contains information to clarify the goal,[2,33,54,69] is delivered by educators with perceived credibility,[11,33,34,98-101] and strategic use of both early and delayed feedback.[87,98] (For more information, see supplementary information: S2.4 'Discussion: Influences on performance due to variations in the constituents of feedback interventions' section).

Influences on performance due to factors beyond feedback

Performance improvement is not solely related to feedback. In our review, other important factors influencing performance, such as instruction and practice

opportunities, also varied between studies. These included teaching and expert demonstration,[75,79,87,98,102-106] learners' background, task complexity and practice opportunities.[5,20,54,107] (For more information, see supplementary information: S2.4 'Discussion: Influences on performance due to factors beyond feedback).

Review limitations

The review has a number of limitations. Despite our attempts to be thorough, we may have missed studies that should have been included. As a number of studies did not report data that would allow easy pooling of data, we either calculated an estimate from available data (including reading off graphs) or excluded the study. Most included studies were conducted in a simulated environment, at Kirkpatrick evaluation level two (change in skills), with only a few situated in authentic clinical practice at Kirkpatrick level three (change in skills applied at work) which may limit application to routine clinical practice.[108]

Implications for future research and clinical practice

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

2.7 Conclusions

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

Supplementary information S2.1

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

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

25	(primary care adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	10661
26	(primary health* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2156
27	(registrar or registrars or senior house officer* or resident or residents or hospital medical officer* or intern or interns or house officer*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	129452
28	dentist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	115221
29	(dent* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	21015
30	nurs*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	629304
31	(midwife or midwives).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	18155

32	(midwife* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	965
33	(allied health adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	12936
34	physiotherapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5348
35	physical therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	5123
36	(physiotherap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	731
37	(physical therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	647
38	occupational therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	4500
39	(occupational therap* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1058
40	speech therap*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7630
41	(speech therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	22
42	speech language therapist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	136

43	(speech language therapy adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	14
44	speech pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	656
45	(speech pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	31
46	speech language pathologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1561
47	(speech language pathology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	105
48	dietician*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1352
49	(dietetic* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	603
50	podiatrist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	626
51	(podiatry adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	19
52	chiropodist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	119
53	(chiropody adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	2
54	social worker*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8079
55	(social work* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	8517
56	psychologist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	11986
57	(psychology adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1532
58	(osteopath* or osteopathic physician*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	6480
59	(osteopath* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	296
60	chiropractor*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	1148
61	(chiropract* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	241
62	pharmacist*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	27316
63	(pharmac* adj2 (staff or personnel or faculty or worker* or provider* or practitioner* or professional* or specialist* or consultant* or student* or trainee* or undergraduate*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7184

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

Supplementary information S2.2

Results

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

Included studies

Participants

Participants included 290 (60%) medical students in four studies,[68,71,75,76] 60 (12%) dental students in one study[70] and 138 (28%) doctors (doctors training in surgery in three studies,[69,72,74] training in obstetrics and gynaecology in one study[73] and training in emergency medicine in one study,[78] and physicians in one study[77]).

Participants were novices to the assessed task in five studies (5/11, 45%);[68-70,75,76] and had prior experience in six studies.[71-74,77,78]

Workplace tasks and Settings

All studies evaluated performance of a discrete task; there were no longitudinal evaluations. The task occurred in simulation settings in seven studies (7/11, 64%) and clinical practice in four studies (4/11, 36%). The task was a surgical procedure in seven studies (7/11, 64%). Five studies involved simulated surgical tasks including bench top models for knot tying[75] and forming a bowel anastomosis;[74] using a laparoscopic simulator for suturing and knot tying;[76] and using a virtual reality (VR) simulator for laparoscopic surgery[68] and endovascular surgery.[69] Two studies involved laparoscopic surgery in clinical practice.[72,73] The remaining four studies evaluated simulated matching of tooth colour in a dental school,[70] simulated cardiopulmonary resuscitation (CPR),[71] chest ultrasound for emergency trauma patients[78] and teaching skills in clinical practice.[77]

Feedback Interventions

The feedback source involved a subject expert in all comparisons except two, including one that compared peer feedback with no feedback,[71] and one that compared expert feedback, peer feedback and no feedback.[69] Feedback occurred while the participant

performed the task (during) in one study,[71] both during and directly afterwards in two studies,[68,76] directly afterwards in four studies,[69,70,73,78] after a delay in three studies[72,74,77] and one study compared feedback during, feedback directly afterwards and no feedback.[75] In addition to evaluative performance information (as per inclusion criteria), the feedback included corrective advice in all studies except one[78] and one where it was unclear.[69] Feedback included additional information from a simulator in three studies,[68,69,76] a video of the participant's performance in two studies[72,77] and written performance information in two studies.[70,77]

Teaching and Practice

In addition, instruction and expert demonstration of the task were provided in six studies (6/11, 55%), including all five studies involving novice participants[68-70,75,76] and one study that involved CPR for medical students, many of whom had previously attended a course.[71] The other five studies involved doctors working in clinical practice; in these studies, no instruction or expert demonstration was included within the research intervention but may or may not have occurred during the course of routine work during that time. One study involved physicians' teaching on ward rounds[77] and the other four studies assessed tasks by doctors training in relevant specialties.[72-74,78]

The amount of practice varied substantially between different studies, for both simple and complex tasks. For example, comparing two studies that involved simple surgical knot tying: in Xeroulis,[75] participants had 18 practice attempts in one session and in O'Connor,[76] they could practice up to an hour a day, for 24 days. Looking at more complex surgical procedures, such as simulated surgery using a virtual reality (VR) simulator: in Ahlborg,[68] participants had two practice attempts at the simulated surgery (laparoscopic salpingectomy) and in Boyle,[69] participants had five attempts at the simulated surgery (renal artery angioplasty and stenting) before the performance evaluation.

Intervention period

The intervention period ranged from one day (most common) up to two months.[72] Nine (9/11, 82%) studies involved a single session (involving one episode of feedback in five

studies[70,73,74,77,78] and multiple episodes of feedback in four studies[68,69,71,75]). Two studies (2/11, 18%) had a longer intervention period involving multiple feedback sessions: one study[72] included approximately four coaching sessions regarding bariatric surgery across a two month surgical attachment, and another[76] included almost daily one hour practice sessions for laparoscopic suturing, with feedback throughout each one, over four weeks.

The timing of the post-feedback performance assessment, in relation to the intervention, differed. It occurred directly following the intervention in seven studies: at the end of the single session in five studies[68,69,71,73,75] and at the end of an extended intervention period in two studies.[72,76] In the other four studies, the post-feedback performance assessment occurred some weeks after the intervention was completed but while relevant exposure to possible teaching and/or practice opportunities continued. Olms[70] included a single feedback session, with the final evaluation two weeks later, in the midst of a routine one month university teaching unit on tooth shade matching. Skeff[77] arranged a single coaching session on ward round teaching in the middle of physicians' four week ward duty, with the final evaluation post-performance evaluation at the end. Soucisse[74] also organised a single coaching session for surgical residents, with the final evaluation occurring three weeks later. Vafaei[78] involved a single workplace-based assessment with feedback for doctors training in emergency medicine on chest ultrasound for emergency trauma patients, followed by a two month period of routine clinical work before the post-feedback assessment.

Research funding

Regarding research funding, one study[71] that focused on cardiopulmonary resuscitation (CPR) quality, was loaned a device (used to measure CPR parameters and provide automated feedback to participants) for the period of the study by Philips but the company was not otherwise involved in the research; five studies received funding from independent institutions,[68,72,73,75,77] three studies did not receive any funding[69,70,78] and two studies did not report information on funding.[74,76]

Risk of bias

Five trials described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'.[70-74] The other six trials simply stated participants were 'randomised' and had no information on allocation concealment, so we rated these studies as 'unclear'. We analysed baseline performance because, although randomisation removes the need to check comparability in baseline task performance for intervention and comparison groups, it may be useful to check this when participant numbers are small and performance improvement is more likely when baseline performance is low.[33] Seven studies reported no statistically significant differences between baseline performances for the comparison groups.[70,73-75,77,78] and four studies did not report baseline task performance.[68,69,71,76] The participants and research team members were not blinded in any included studies because the intervention involved feedback between a research team member and a participant, consistent with most education interventions. However, in all included studies, we thought this was not likely to influence the outcome (post-intervention performance assessment) because implementation and adherence to the intervention were not affected. In eight studies the outcome was assessed by either blinded assessors who rated videos of the participants' performance[69,72-75,77] or by a machine (simulator or CPR machine),[68,71] so we rated these as 'low risk' of bias. In three studies, the feedback provider and outcome assessor appeared to be the same person, so these were rated as 'high risk'.[70,76,78] Across all the studies, the follow up rate for each group was at least 85%. Only two studies had a prior published protocol in addition to reporting all outcomes as planned.[72,74] For all other studies, it could not be ascertained if outcomes had been selectively reported, so these were rated as 'unclear', except one. This one study was rated as 'high risk' for selective outcome reporting because it did not include the expected information on performance post-intervention.[76]

In summarising the risk of bias across domains within each study, two studies had all domains rated 'low risk, so these were rated low risk.[72,74] Six studies had at least one domain with 'unclear' risk but no 'high risk' ratings, so these were rated as 'unclear' risk

of bias.[68,69,71,73,75,77] Three studies had at least one domain at high risk of bias, so we judged these studies to be at 'high risk' of bias.[70,76,78]

Certainty of evidence

For the comparison of verbal face-to-face feedback compared to no feedback, excluding studies at high risk of bias, we graded the quality of evidence for the outcome of 'objective assessment of a health professional's performance'. The risk of bias was rated as 'unclear' across multiple included studies and the overall body of evidence indicated this was likely to seriously alter the results, so we downgraded the overall evidence by one level. The two aspects that were most influential on our decision were the lack of allocation concealment and prior published protocols to preclude selective reporting of outcomes. Participant and research team member blinding was not possible due to the intervention. However, this had limited impact on the selected outcome 'objective assessment of performance', as no changes occurred in intervention implementation or adherence as a consequence of this lack of blinding.[60] We judged the results to be directly applicable to our review question and therefore the evidence was not downgraded for indirectness. There was some methodological and statistical heterogeneity across studies (the test for heterogeneity was not significant with $P = 0.14$ and $I^2 = 34\%$), which was not explained by subgroup analysis. However, all studies reported a beneficial effect, so the uncertainty seemed to lay in the magnitude of effect rather than the presence of an effect. Therefore, we decided not to downgrade the evidence due to inconsistency.[80] We judged the effect size to be sufficiently precise and therefore did not downgrade the evidence for imprecision of results. This was based on sufficient numbers of participants (392 when studies with high risk of bias were excluded) and a consistent beneficial effect, indicated by the confidence interval for the overall effect estimate not crossing zero and all individual studies showing a beneficial effect with substantial overlap in their confidence intervals. Finally, we judged that there was likely to be a systematic overestimation of the underlying beneficial effect of feedback because we strongly suspected publication bias (see Funnel plot 5b) and therefore we downgraded the evidence by one level.

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

Supplementary information S2.3

Results

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

Included studies

Participants

Included studies involved medical students (1076, 55%) in 14 studies,[71,75,76,81,83,84,87-94] mixed health professional students (640, 32%) in one study,[82] pharmacy students (153, 8%) in one study[86] and doctors (105, 5%) in four studies.[69,77,85,95]

Participants were novice to the task in 11 studies (11/20, 55%). Three studies documented prior experience: one study involved attending physicians teaching on ward rounds with a range of experience[77] and two studies documented previous training including CPR[71] and history taking and communication skills in medical students.[88] The remaining six studies did not report this information. One of these studies[82] involved teaching CPR to first year health professional students across a mix of disciplines, some of whom may have had prior experience. One study[85] involved evaluating professional and clinical skills in first year paediatric residents who likely had relevant training as medical students. In two of these studies, the baseline performance of junior medical students' attempting surgical knot tying was poor, which suggest limited prior experience.[83,92] In the last two studies there was no information on prior experience: one assessed a simulated medication consultation by third year pharmacy students[86] and another[94] assessed case presentation skills in third year medical students in their paediatric attachment.

Workplace tasks and Settings

All studies included assessment of a discrete task except two studies which involved longitudinal evaluations.[77,85] Three studies evaluated performance in a clinical practice setting (involving teaching skills[77] professional and communication skills[85] and oral case presentations[94]) and the remaining 17 assessed performance in a simulated

environment.[69,71,75,76,81-84,86-93,95] Simulated surgical tasks included suturing and/or knot tying,[75,76,81,83,92] bench top surgical procedures such as vascular anastomosis,[95] flexible ureteroscopy for urolithiasis,[87] renal artery angioplasty and stent placment,[69] or surgery using a VR simulator for a laparoscopic salpingectomy.[89] Simulated critical care tasks included basic life support (BLS)/CPR,[71,82] intubation[93] and pharmacist-patient consultation.[86] The remaining simulated tasks included a hearing test,[90,91] simulated patient consultation[88] and nasogastric tube insertion.[84]

Interventions

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

Some studies investigated straightforward variations in feedback, including differences in frequency (low or high[84]), stage of practice (early or late[87]), different feedback models ('learning conversation' compared with 'feedback sandwich' frameworks[82]), source expertise (expert or peer[69]) and expert feedback compared to compliments.[83,92] Another collection of studies explored the effect of adding expert feedback to other interventions, such as in addition to simulator performance data[71,89] or to written feedback;[77,85] or adding expert review of a participant's performance video to a practice session in which expert feedback was available.[95] One study[81] compared verbal feedback by an expert who had just directly observed the performance, with written feedback emailed later that day by another expert who watched a video of the performance. Other studies explored more complex phenomena. One study[93] compared two feedback variations in different combinations across four groups. One variation compared an evaluative verbal comment from an expert, to a written numerical performance rating. The second variation involved an individual comparing their performance evaluation to either their own previous attempts (highlighting individual progress) or to expected performance at student, resident or specialist level (comparison with others). Another research group investigated two complex influences in separate studies. One study[91] examined how the credibility of the feedback provider (high or low credibility) influenced learner outcomes. The other study[90] examined the effect of phrasing

corrective information in different ways, so in one intervention corrective information was framed within a positive phrase whereas in the other, it was framed within a critical phrase.

Research funding

One study[71] was loaned a device by Philips as detailed earlier, seven studies received funding from independent institutions,[75,77,84,85,89,93,94] six studies did not receive any funding[69,83,87,88,90,91] and six studies did not report information on funding.[76,81,82,86,92,95]

Risk of bias

The risk of bias assessment for the comparison of verbal face-to-face feedback to alternative feedback is presented in Figure 3. Seven described an adequate method for randomised sequence generation and allocation concealment, so we rated these studies as 'low risk'.[71,81,82,89-91,94] Two studies had adequate random sequence generation, which we rated 'low risk' but had insufficient information on allocation concealment, which we rated 'unclear risk'.[85,88] The remaining studies simply stated participants were 'randomised' and had insufficient information on allocation concealment, both of which we rated 'unclear risk'. Two studies described inconsistencies with randomisation, so these were rated 'high risk' of bias for sequence generation and allocation concealment.[86,93] There was unequal baseline performance between groups reported in one study[91] and identified from another study's data (obtained from authors).[81] No statistically significant differences in baseline performance between groups were reported in seven studies [75,77,83-85,87,91] and baseline performance was not reported in eleven studies.[69,71,76,82,86,88,89,92-95] None could blind participants or research team members due to the face-to-face feedback interventions. However we thought this was not likely to influence the outcome as implementation and adherence to the intervention were not affected in all studies, which were rated 'low risk', except one in which some participants may not have experienced the intervention they were allocated to, so it was rated 'unclear'.[94] The outcome was assessed by blinded assessors or machines in all studies, which were rated 'low risk' except two studies that did not explicitly describe blinded assessors, which were rated 'unclear'[77,91] and four studies

that seemed to have assessors who were aware of participant allocation, so these were rated 'high risk'.[76,86,88] All had high proportions of participant completion data except three[90,91,95] and one report provided insufficient information.[92] Three studies had prior published protocols and reported all outcomes as planned, so they were rated 'low risk'.[82,85,89] All of the others did not have a prior published protocol but did present outcomes as expected and were rated as 'unclear',[69,71,75,77,81,83,84,86-88,90-95] except one study which was rated as 'high risk'.[76]

In summarising the risk of bias across domains within each study, two studies were rated as low risk[82,89] as all domains were rated as 'low risk' of bias, seven studies were rated as 'high risk' because at least one domain was rated as 'high risk',[76,86,90,91,93,95] and the remaining studies were rated as 'unclear' as they had at least one domain with 'unclear' risk but no 'high risk' ratings.

Supplementary material S2.4

Discussion

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

Included studies involved health professional students and clinicians (mainly medical) performing a range of workplace tasks, particularly surgical and most commonly in a simulated environment. The meta-analysis results are dominated by one study,[71] evaluating effective compressions during CPR, which contributed the largest number of participants from a single study to the meta-analysis and had an individual study SMD of 0.25 (95% CI -0.02, 0.51). Several factors may have contributed to the relatively small overall benefit from this feedback intervention compared to many of the other included studies. These include a short practice period with feedback from a peer (as opposed to an expert) who was concurrently performing a different task (the student performing ventilation provided advice on correcting compressions to the student performing compressions).

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

Previous research has identified that feedback is more effective when the content includes information that makes the goal clear (for example, describing correct performance or providing an expert demonstration of the task) and advice on how to improve.[2,33,54] However, detailed specifications about feedback content were often not clearly reported in included studies, which suggests that researchers may not have realised the importance of this. The feedback source was more often experts than peers, in our included studies. One small study[69] directly compared expert feedback to peer feedback for novices learning a surgical task using a visual reality simulator. It did not find a statistically significant difference (SMD 0.46, 95% CI -0.7, 1.61), although there was some indication that learners in the expert feedback group improved faster and their performance was smoother. In earlier systematic reviews[33,34] and other research,[11,98-101] feedback from a highly credible source (expert feedback) has been reported to be more effective. Also, the timing of feedback in included studies varied; it

was provided while the learner undertook the task, immediately afterwards or some time afterwards. One small study,[75] in which novices learnt to suture, feedback during the task was compared to feedback immediately after each attempt. It did not find a statistically significant difference in performance after one hour of practice but did a month later (beyond the scope of our review), in favour of feedback immediately after practice. In another study, in which students practised simulated laparoscopic surgery, the effect of additional expert feedback was compared with performance information provided by the simulator alone. The authors reported that more participants in the 'simulator feedback only' stopped practising. Previous research has noted that for novices learning a complex task, early feedback and assistance may prevent extreme frustration and giving up.[98] Feedback during task performance results in faster initial skill acquisition compared to feedback after task performance, particularly for procedural skills, as errors are corrected in real-time, but poorer subsequent independent performance.[98,103] It is thought that a learner develops a mental schema depicting how to do the task, which they develop during practice attempts and this is utilised for subsequent performances.[104,105] However feedback during task performance appears to interfere with this process, possibly due to cognitive overload.[87,106]

In the second analysis, verbal face-to-face feedback was compared to a multitude of alternative feedback interventions. In addition to feedback source, frequency, timing and content, there were differences across feedback modality (verbal, written, numerical, video or machine output information), feedback format (coaching, 'learning conversation' or 'feedback sandwich'), phrasing of feedback (expressing the same corrective information in a positively or negatively couched phrase), benchmarks set for learners (comparing current performance with previous own scores or training level benchmarks) and feedback compared to compliments. Each study discussed and revealed useful insights into components that might influence feedback outcomes.

Influences on performance due to factors beyond feedback

Firstly, teaching and expert demonstration were common (but not standard) and the amount and type varied across studies, which have previously been shown to impact

performance.[79,102] Practice opportunities also differed enormously across included studies, even those involving similar tasks. In addition, there was variation across learners' prior relevant expertise (e.g. first year medical students or surgical trainees learning a surgical task) and the complexity of the task (knot tying or laparoscopic bariatric surgery). Previous research has shown that teaching and expert demonstration assist a learner to improve, particularly in the initial phases of skill acquisition[79,102] and practise is essential for mastering any skill.[5,20] Furthermore, learners who are motivated to learn a challenging but achievable skill are most likely to improve their performance, according to 'goal setting' and 'self-determination' theories.[54,107]

2.8 Chapter 2 Closing discussion

To our knowledge, this systematic review is the first to confirm that verbal face-to-face feedback involving health professionals is likely to substantially improve clinical task performance, compared to no feedback (pooled effect size of 0.7; 95% CI 0.37 to 1.03; $p<0.001$). Analysis of the included studies revealed a variety of components within face-to-face feedback, challenging the notion of feedback as a single entity and highlighting its multifaceted nature. The range of effects suggested that certain feedback components, or combinations, (or other factors, such as the task being taught), may exert different impacts on performance; at best, this could markedly enhance performance indicated by an effect size of approximately one at the upper confidence limit. Many studies did not clearly describe the various potential influences on performance within the intervention, suggesting that authors may not have recognised a role for different elements within the feedback interaction in shaping outcomes. The framework used to present the key characteristics of the included studies (presented in Tables 2.1 and 2.2 in the article) may assist future research design and reporting in highlighting aspects of feedback that could be included in the feedback process. The feedback interventions within included 'alternative feedback' studies were markedly diverse, underscoring the potential assortment of influences under the umbrella term of 'feedback'. However, the quality of many of the included studies was low. This was primarily due to risk of bias, frequently related to a lack of prior publication of the protocol (particularly for studies published before such transparency became a standard for trial reporting), providing insufficient details to confirm random sequence generation and allocation concealment, the inevitable failure to blind participants or study personnel to group allocation during a face-to-face feedback intervention and publication bias.

These limitations not withstanding, establishing what could be known about empirical support for the potential benefits of feedback discussions provided the impetus for developing of an instrument that could distinguish and clearly characterise the multifaceted components of effective feedback in clinical practice.

Chapter 3

Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement

3.1 Chapter 3 Introduction

In Phase 1, the systematic review reported in Chapter 2 provided empirical support for the likely value of verbal face-to-face feedback and underscored the key point that feedback incorporates multiple components with differing effects.

Based on this, Phase 2 of this research, targeted the development of a comprehensive set of descriptions of educator behaviours designed to optimise learner outcomes, constituting a provisional feedback evaluation instrument. Phase 2 comprised two distinct bodies of work, described in stages. Phase 2, Stage 1 involved another review of the literature, this time searching for what was known about the constituents of quality feedback practice. In Phase 2, Stage 2, these constituents were expressed as observable behaviours and revised in collaboration with a team of experts, using Delphi methodology.[109] This resulted in consensus on a parsimonious set of educator behaviours that exemplified high quality feedback practice.

This research, directed at developing the Feedback Quality Instrument, focused on the educator's role during verbal face-to-face feedback following observation of a learner's performance in clinical practice. Optimising feedback discussions requires collaboration between the educator and learner, but each plays a different role. The educator has more power and responsibility to helpfully influence the interaction. Additionally, this approach reflected my own desire to understand, "How can I do this well?" as the initial driver for the research. Despite this focus on the educator's perspective, the desired

outcome is centred on benefiting the learner. There is no question in my mind that the learner's role is equally important, but it is not the focus of this work.

Phase 2, Stage 1 involved a second literature review. While the systematic review reported in Chapter 1 was essential in understanding what was known about the effects of feedback, it provided little guidance regarding an educator's role in promoting quality feedback practice. The aim of this research was to identify discrete elements of an educator's role that could influence learner outcomes and were substantiated by research findings rather than being based solely on expert opinion. Collating this information was challenging, because information on best practice in feedback was distributed across many fields of research and it could not be identified efficiently using routine searching methods. At the time, there was no specific Medical Subject Headings (MeSH) terms for human task-performance feedback and the indexed term included all types of feedback including biological. Attempts to search using typical search strategies led to vast and poorly targeted results. Consequently, a snowballing methodology was employed. In this approach, key literature, deemed to be valuable reference works in the field of feedback, was used as the starting point for searching. Using these articles and book chapters, whenever experts made recommendations regarding quality feedback practice, the citations provided to support them were investigated to identify any substantiating research. In addition, reference lists were searched. The literature was also searched for feedback instruments that had been designed for face-to-face verbal feedback interactions in health professions education, to check whether a suitable feedback instrument already existed, or to provide items to consider in the development of such an instrument. In this way, the literature was explored up to the point where little new information was appearing, and a collection of relevant literature was assembled. Information from this broad literature, across multiple disciplines, was distilled to catalogue key elements of an educator's role considered to be important in effective feedback.

Phase 2, Stage 2 involved crafting these elements so that they described observable behaviours, and refining them into a parsimonious set of statements that clearly

described important educator behaviours in quality feedback interactions. To assist with the process of refinement, a team of experts was assembled. A Delphi process was selected as the best way to harness the opinions of these experts. Using online survey software (SurveyMonkey Inc., San Mateo, California, USA. www.surveymonkey.com) the initial set of observable behaviours, developed by the core research team (CEJ, JLK, EKM), was distributed to each member of the panel. The experts individually rated the importance and phrasing of each educator behaviour, and had the opportunity to add comments or suggest revisions. Then the descriptions were revised based on this input. In subsequent rounds, a link to a new survey containing the revised set of observable behaviours was circulated, plus the ratings and comments from the previous round and a document that linked each revised items to the previous versions. In addition, the material was individually designed for each panel member, by including their own ratings and comments from the previous round so that each participant could compare their own opinion regarding the pooled and deidentified comments made by the panel as a whole. Participating experts were enthusiastic and committed to advancing feedback practice, and the strong diversity and expertise within the group fostered rich dialogue. The advantage of using a Delphi technique was that it allowed each expert to individually consider the questions whilst also having the opportunity to reflect on opinions from the rest of the group. This allowed everyone's perspective to be aired in a balanced way and promoted consensus amongst the group, as items evolved. It also reduced the risk of the most prominent or extroverted panel member dominating the discourse, as can occur in focus groups, which could have been used as an alternative method. In addition the on-line nature of the Delphi technique enabled panel members to complete their review and feedback at times that were convenient to each of them, in contrast to the difficulties associated with bringing experts together at a single time and location.[110] Nevertheless, such an ambitious step was undertaken in the final stages of item resolution where face-to-face discussion was successful in achieving consensus on the final item set and item descriptions.

Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement

This chapter is the study published in BMC Medical Education DOI: 10.1186/s12909-016-0613-5 (Lightly edited for headings and formatting consistency). Published March 2016.

Christina E. Johnson[1;] Jennifer L. Keating[2;] David J. Boud[3]; Megan Dalton[4]; Debra Kiegaldie[5]; Margaret Hay[1]; Barry McGrath[6]; Wendy A. McKenzie[7]; Kichu Balakrishnan R. Nair[8]; Debra Nestel[1]; Claire Palermo[9]; Elizabeth K. Molloy[1]

[1] Health Professions Education and Educational Research (HealthPEER), Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[2] Department of Physiotherapy, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[3] Faculty of Arts and Social Sciences, University of Technology Sydney, Ultimo, Australia;

Centre for Research on Assessment and Digital Learning, Deakin University, Geelong, Melbourne; Institute of Work-Based Learning, Middlesex University, London, UK.

[4] School of Human, Health and Social Science, Central Queensland University, Rockhampton, Australia

[5] Faculty of Health Science, Youth and Community Studies, Holmesglen Institute and Healthscope Hospitals, Holmesglen, Melbourne, Australia

[6] Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[7] Faculty of Education, Monash University, Melbourne, Australia

[8] School of Medicine and Public Health, Faculty of Health and Medicine, University of Newcastle, Newcastle, Australia

[9] Department of Nutrition and Dietetics, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

Keywords: Feedback, Clinical practice, Delphi process, Health professions education, Educator behaviour

Funding: Christina Johnson, Jennifer Keating and Elizabeth Molloy received the following grant:

Office of Learning and Teaching (OLT) and Higher Education and Research Development Society in Australasia (HERSDA) Researching New Directions in Learning and Teaching: Seed Funding to Support Project Mentoring and Collaboration.

Declaration of interest: None

3.2 Abstract

Background: Health professions education is characterised by work-based learning and relies on effective verbal feedback. However the literature reports problems in feedback practice, including lack of both learner engagement and explicit strategies for improving performance. It is not clear what constitutes high quality, learner-centred feedback or how educators can promote it. We hoped to enhance feedback in clinical practice by distinguishing the elements of an educator's role in feedback considered to influence learner outcomes, then develop descriptions of observable educator behaviours that exemplify them.

Methods: An extensive literature review was conducted to identify i) information substantiating specific components of an educator's role in feedback asserted to have an important influence on learner outcomes and ii) verbal feedback instruments in health professions education that may describe important educator activities in effective feedback. This information was used to construct a list of elements thought to be important in effective feedback. Based on these elements, descriptions of observable educator behaviours that represent effective feedback were developed and refined during three rounds of a Delphi process and a face-to-face meeting with experts across the health professions and education.

Results: The review identified more than 170 relevant articles (involving health professions, education, psychology and business literature) and 10 verbal feedback instruments in health professions education (plus modified versions). Eighteen distinct elements of an educator's role in effective feedback were delineated. Twenty five descriptions of educator behaviours that align with the elements were ratified by the expert panel.

Conclusions: This research clarifies the distinct elements of an educator's role in feedback considered to enhance learner outcomes. The corresponding set of observable educator behaviours aim to describe how an educator could engage, motivate and

enable a learner to improve. This creates the foundation for developing a method to systematically evaluate the impact of verbal feedback on learner performance.

3.3 Background

Health professions education is characterised by work-based learning where a student or junior clinician (a 'learner') learns from a senior clinician (an 'educator') through processes of modelling, explicit teaching, task repetition, and performance feedback.[21,22] Feedback, which follows an educator observing a learner perform a clinical task, is an integral part of this education. This may occur 'on the run', during routine clinical practice or as scheduled feedback during workplace-based assessments, planned review sessions, or at mid- or end-of-attachment performance appraisals.

Feedback has been defined as a process in which learners seek to find out more about the similarities and differences between their performance and the target performance, so they can improve their work.[19] This definition focuses on the active role of the learner and highlights that feedback should impact on subsequent learner performance.

Feedback needs to help the learner develop a clear understanding of the target performance, how it differs from their current performance and what they can do to close the gap.[1,6,18] To accomplish this, a learner has to construct new understandings, and develop effective strategies to improve their performance. A learner also has to be motivated to devote their time and effort to implementing these plans, and to persist until they achieve the target performance.

In an attempt to enhance learner-centred feedback, it is enticing to focus on the learner and their role in the feedback exchange. However given that educators typically lead educational interactions, particularly in the early stages, targeting the educator's role in feedback may have a greater influence in cultivating learner-centred feedback. A skilled educator can create an optimal learning environment that engages, motivates and supports learners, thereby enabling them to take an active role in evaluating their performance, setting valuable goals and devising effective strategies to improve their performance.[29,111] Learners who have experienced such sessions could then carry forward a clear model of high quality feedback into future interactions throughout their professional life.

Experts in health professions education assert that feedback is a key element in developing expertise.[3-6,50,112,113] Learners in the health professions also believe feedback can help them and they want it.[7-9,36] However there is limited evidence to support this conviction that feedback improves the performance of health professionals. The strongest evidence is from two meta-analyses, which indicated that audit followed by feedback improved adherence to clinical guidelines.[33,34] Beyond the health professions there is stronger evidence. In a synthesis of 500 meta-analyses (180,000 studies), feedback was reported to have one of the most powerful influences on learning and achievement in schools.[1] Another meta-analysis of 131 studies compared feedback alone with no feedback on objective measures of performance of diverse tasks. That analysis also supported the conclusion that feedback improved performance.[2]

Despite the enviable theoretical benefits of feedback, problems have been reported in practice. In observational studies of face-to-face feedback, educators often delivered a monologue of their conclusions and recommendations. Learners spoke little, asked few questions, minimised self-assessment (if asked) and were not involved in deciding what was talked about, explaining their perspective or planning ways to improve.[38,114-118]

Observational studies and reviews of feedback forms indicated that educators' comments were often not specific, did not identify what was done satisfactorily and what needed improvement, and did not include an improvement plan.[114,119-121]

Educators have reported that they did not feel confident in their feedback skills. In particular they avoided direct corrective comments as they feared it could undermine a learner's self-esteem, trigger a defensive emotional response or spoil the learner-educator relationship. Educators experienced negative feelings themselves, such as feeling uncomfortable or mean.[12,36,38,114]

Feedback does not always improve performance and can even cause harm.[1,33,34,100,122] In Kluger and DeNisi's meta-analysis, approximately a third of studies found that performance deteriorated following feedback.[2]

Learners have reported that they do not always implement feedback advice. Their reasons included they did not consider there was a problem, did not believe the educator's comments were credible or relevant, [11,13] or did not understand what needed improving or how to do it.[11,37] Learners have also reported experiencing strong negative emotions such as anger, anxiety, shame, frustration and demotivation following feedback, especially if they thought feedback comments were unfair, derogatory, personal or unhelpful.[35-37,123]

Our goal is to promote high quality feedback by helping educators to refine the way they participate in feedback, and subsequently to enhance learner outcomes. It is not clear what comprises high quality, learner-centred feedback or how educators can promote it.[124,125] One explanation for the mismatch between the theoretical benefits of feedback and the problems experienced in practice is that feedback involves multiple unidentified elements that may influence the outcome. Therefore it would be useful to clarify the components of an educator's role in feedback required to achieve the aim of engaging, motivating and enabling a learner to improve their skills and develop a list of key educator behaviours that describe how those objectives could be accomplished in clinical practice.

We chose to target the educator's role first because educators have substantial influence and have a primary responsibility to model high quality feedback skills. The setting we focused on was scheduled face-to-face verbal feedback following observation of a learner performing a task, as this is a particularly common form of feedback in the workplace education of health professionals.

3.4 Methods

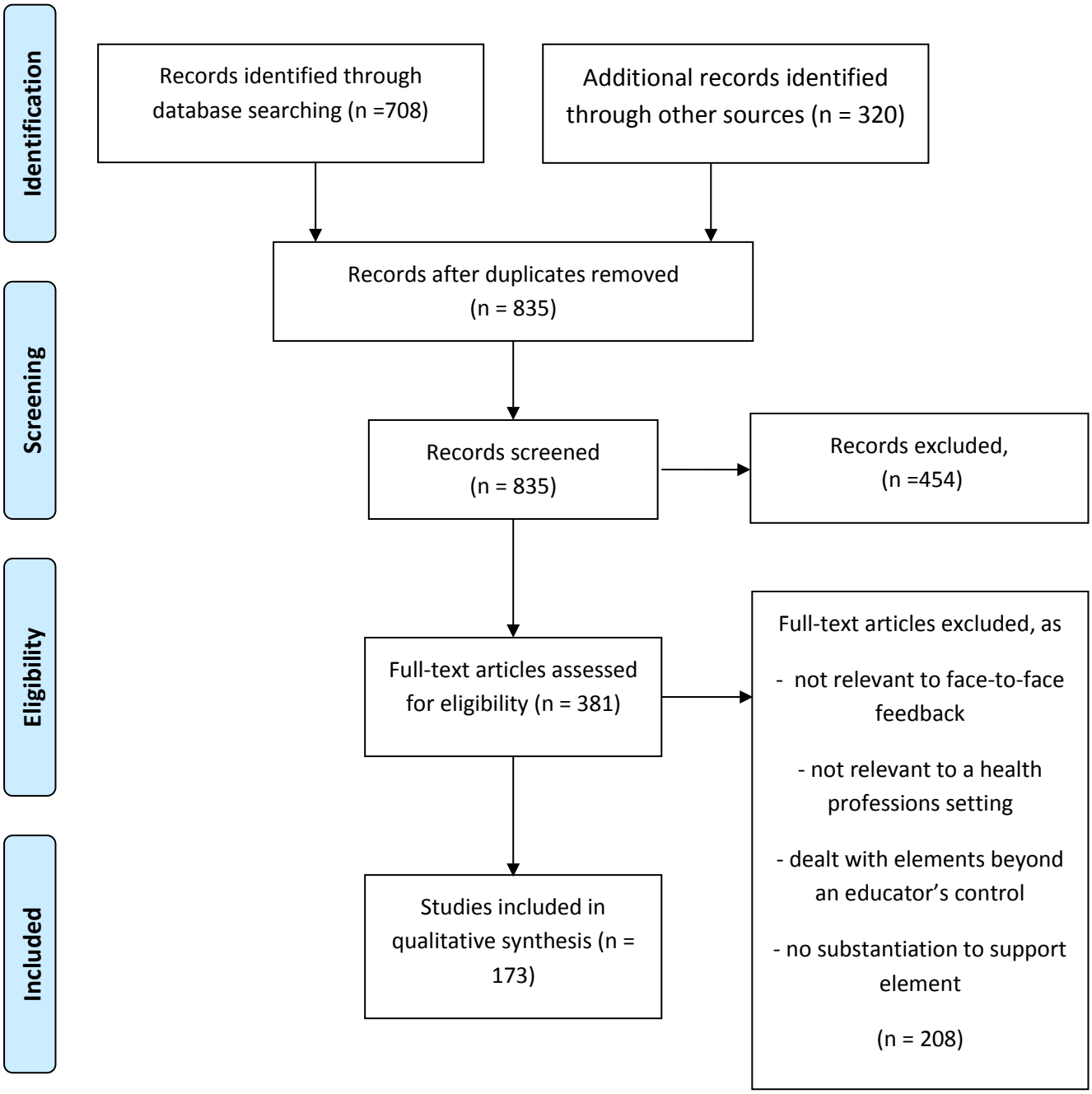
In this paper we describe the first phase in this process, which had two stages. The first stage involved conducting an extensive literature review to delineate the key elements of an educator's role in effective feedback. In the second stage, a set of correlated educator behaviours was created and then refined in collaboration with an expert panel.

Stage 1: Literature Review

The literature review was conducted to identify distinct elements of an educator's role in feedback asserted to help a learner to improve their performance and the supporting evidence. The elements describe the key goals of an educator in high quality, learner centred feedback i.e., what needs to be achieved but not necessarily how to do it. In addition published instruments (or portion thereof) designed to assess face-to-face verbal feedback in health professions education were reviewed for descriptions of educator behaviours considered to be important in effective feedback.

The target information was embedded within diverse articles spread across a broad literature base and was poorly identified by standardised database search terms. We therefore utilised a 'snowball' technique.[50,126] This began with identifying systematic reviews on feedback plus published articles and book chapters in the health professions, education, psychology and business by prominent experts. When authors cited articles to support claims and recommendations, the original substantiating source was traced. Additional relevant articles were identified through bibliographies and citation tracking. This continued to the point of saturation where no new elements were identified. In addition, published instruments (or portion thereof) designed to assess face-to-face verbal feedback in health professions education were searched to identify relevant educator activities. Published literature was searched across the full holdings of Medline, Embase, CINAHL, PsycINFO and ERIC up to March 2013, and then continued to be scanned for previously unidentified elements until September 2015 (see Figure 3.1).

Figure 3.1: PRISMA 2009 Flow Diagram



Element construction

Elements were constructed by analysing and triangulating supporting information extracted during the literature review, including items in feedback assessment instruments. Potential elements and the substantiating evidence or argument were extracted by one researcher (CEJ) and verified by core research team members (JLK and EKM). Similar elements were grouped and those with overlapping properties were collapsed. The core research team used an iterative process of thematic analysis to develop a list of elements that described distinct aspects of an educator's role in feedback.[127]

Stage 2: Development and refinement of the educator behaviour statements

The next step was to operationalise the elements by reconstructing them as statements describing observable educator behaviours that exemplify high quality feedback in clinical practice. An initial set of statements was developed by the core research team, using the same iterative process of thematic analysis, in accordance with the following criteria: the statement describes an observable educator behaviour that is considered important for effective feedback that results in improved learner performance; targets a single, distinct concept; and uses unambiguous language with self-evident meaning.[128]

A Delphi technique was used to develop expert consensus on the statement set, in which sharing of anonymous survey responses enables consensus to develop as opinions converge over sequential rounds.[109,110,129] An expert panel was formed. All panel members provided informed consent. Members refined the individual statements and the composition of the list as a whole, and developed consensus on each statement (defined as over 70% panel agreement) during three rounds using a Delphi technique.[130]

Expert Panel

The research team invited nine Australian experts with experience in health professions education, feedback, psychology, education and instrument development to join research team members (JLK and EKM) to create a panel to refine the statement set. The primary researcher (CEJ) acted as the facilitator. A structured survey presenting the

initial statements was distributed to panel members using online survey software. For each statement, panel members were asked to consider two questions i) importance: 'this statement represents an important educator behaviour in verbal feedback' (rating options were 'very unimportant, unimportant, neutral, important, very important or don't know') and ii) phrasing: 'this statement meets the specified criteria' (rating options were 'agree, neutral, agree, strongly agree or don't know'). For each question, panel members were asked to provide their reasoning and additional comments in free text boxes. Criteria for each statement and examples of two questions from the survey are presented in Figure 3.2.

After each round, the ratings and comments were analysed using an iterative process of thematic analysis, and the educator behaviour statements refined accordingly.[127] For the following round, a revised set of statements was circulated. This was accompanied by summarised anonymous panel responses from the previous round for participants to consider before continuing with the survey.

Following the conclusion of the three Delphi rounds, a face-to-face meeting of panel members was convened to resolve outstanding decisions. The meeting was audiotaped, transcribed verbatim and analysed using thematic analysis, and a set of educator behaviours was finalised. Ethics approval for this study was obtained from Monash University Human Research Ethics Committee Project Number: CF13/1912-2013001005.

Figure 3.2: Desirable criteria and example of two questions from Delphi Round 3 survey

Delphi Round 3

The importance of Statement 1:

'The educator's comments were based on directly observed performance'

	very unimportant	unimportant	neutral	important	Very important	don't know
This statement represents an important educator behaviour in verbal feedback	O	O	O	O	O	O

Please give your reasoning and additional comments

Phrasing of Statement 1:

'The educator's comments were based on directly observed performance'

	strongly disagree	disagree	neutral	agree	strongly agree	don't know
This statement meets the specified desirable criteria	O	O	O	O	O	O

Suggest alternative phrase or give additional comments

Desirable criteria for each statement:

Each statement should

- be important for effective feedback that results in improved learner performance
- describe an observable behaviour
- target a single aspect of feedback only
- be unambiguous

3.5 Results

Literature Review

The database search identified a key set of reports that led to the identification of more than 170 relevant articles.[1-3,24,33,34,50,112,131-136] These articles included observational studies of feedback, interviews and surveys of educators and learners, summaries of written feedback forms, feedback models, eminent expert commentary, consensus documents, systematic reviews and meta-analyses, and established theories across education, health professions education, psychology and business literature. There was little high quality evidence to clarify the effects of specific elements of feedback.

Literature Review: Elements

Eighteen elements that describe the educator's role in high quality feedback, were created by identifying substantiating information offered to support expert argument across diverse literature. These are presented in Figure 3.3. The order is aligned to the usual flow of a feedback interaction including set up, discussing the assessment and developing an action plan (including some elements that apply throughout).

Literature Review: Face-to-face verbal feedback instruments

The literature search identified 10 instruments (and additional modified versions) that, to some extent, assessed face-to-face verbal feedback in health professions education. It was hoped that these instruments would include items that described educator behaviours associated with effective feedback in clinical practice. However none of these instruments were designed to assess an educator's contribution to an episode of face-to-face verbal feedback following observation of a learner performing a task in the workplace. Three instruments assessed a simulated patient's feedback comments,[137-141] three assessed an instructor's debriefing to a group following a healthcare simulation scenario[142-144] two instruments assessed brief feedback associated with an Objective Structured Clinical Examination (in which the primary aim of the study was to determine if a senior medical student's feedback was of a similar standard to a doctor's),[145,146] and

two longitudinally assessed an educator's overall clinical supervision skills, including feedback, across a clinical attachment.[147-149]

Figure 3.3: Key elements of an educator's role in effective feedback, extracted from the literature

Set Up

1. Educator comments are based on first hand observations of the learner performing a task
2. Educator offers feedback as soon as possible after observation of the task
3. Educator explains the purpose of feedback is to help the learner improve their performance
4. Educator establishes an effective learning environment (non-judgemental, honest, supportive, respectful, trustworthy)
5. Educator describes the proposed process, so the learner knows what to expect
6. Educator promotes learner-centred feedback and encourages learner participation through dialogue
7. Educator collaborates with the learner
8. Educator deals with learner emotions and does not ignore them

Assessment

9. Educator understands the value of and encourages a learner's honest self-assessment
10. Educator clarifies the expected standards and the characteristics of the target performance
11. Educator assessment identifies similarities and differences between learner's performance and the expected standards, so the performance gap and focus for improvement is clear
12. Educator clearly describes a specific incident, explains their concern and explores the learner's perspective to identify the basis for the learner's actions and any learning needs.

13. Educator comments are specific (detailed and related to a specific incident) and accompanied by the reason, so it is useful for the learner
14. Educator comments relate to the task not personal characteristics of the learner
15. Educator helps the learner to prioritise topics that are most useful (important and relevant) to discuss

Action Plan

16. Educator helps the learner to set goals that maximise performance
17. Educator helps the learner to develop a learner-tailored action plan involving ideas for improvement and practical steps to achieve the learning goals
18. Educator plans with the learner to review the impact of feedback on subsequent performance

Development and refinement of the educator behaviour statements using a Delphi technique

Panel

All nine invited experts agreed to participate to create an eleven member panel; the primary researcher acted as facilitator. All panel members had senior education appointments at a hospital or university (the majority were professors and/or directors). The panel included seven health professionals (medicine, nursing, physiotherapy, dietetics and psychology) and several internationally recognised experts in feedback, education and training, simulation and instrument development. There was a high level of engagement by the panel throughout; all members completed each survey in full and made frequent, detailed additional comments.

Development of observable behaviour statements

The initial set of observable educator behaviours, developed by the core research team from the elements, contained 23 statements as some elements required more than one for operationalisation. This set was submitted to the Delphi process. After every round,

the individual statements and the set as a whole were modified, based on the panel's ratings and comments. Revisions included refining statements to better target the underlying concept, and rewording statements to better align with the specified criteria (see Figure 3.2). Overlapping statements were combined and new ones were developed.

One example of how an element was refashioned into a corresponding observable educator behaviour, is described here. Element 4 states an "educator establishes an effective learning environment". This was operationalised into "the educator showed respect and support for the learner" (Behaviour Statement 11) and "the educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise" (Behaviour Statement 4).

After completion of the third round, there were 25 statements in the set. Expert consensus was achieved for i) statement importance: all except one and ii) statement phrasing: all except three. These outstanding issues were resolved at the face-to-face panel meeting. The final list, presented in Figure 3.4, included 25 statements that explicitly describe observable educator behaviour in high quality verbal feedback.

Figure 3.4: List of educator behaviours that demonstrate high quality verbal feedback in clinical practice.

1. The educator's comments were based on observed performance
2. The educator offered to discuss the performance as soon as practicable
3. The educator explained that the purpose of feedback is to help the learner improve their performance
4. The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise
5. The educator described the intended process for the feedback discussion
6. The educator encouraged the learner to engage in interactive discussions
7. The educator asked the learner about their learning priorities for the observation and feedback discussion, and responded to them
8. The educator encouraged the learner to consider the issues and possible solutions during the feedback discussion
9. The educator encouraged the learner to discuss difficulties and ask questions regarding the performance so the educator could help the learner to develop solutions
10. The educator acknowledged and responded appropriately to emotions expressed by the learner
11. The educator showed respect and support for the learner.
12. The educator asked what the learner understood about the benefits of self-assessment and helped clarify
13. The educator asked the learner to identify key similarities and differences between the learner's performance and the target performance
14. The educator clarified with the learner key features of the target performance and explained the reasoning
15. The educator clarified with the learner similarities and differences between the learner's performance and the target performance
16. The educator's comments focused on key issues for improving the performance
17. First the educator described, using neutral language, what the learner did (action, decision or behaviour), and the consequences
18. The educator clearly explained their perspective on the learner's actions, including the reason for their concern
19. The educator explored the learner's perspective and reasoning to reveal the basis for the learner's actions (e.g. what was the learner trying to do and options considered/ difficulties encountered)

20. The educator's comments were focused on the learner's actions, not personal characteristics
21. The educator helped the learner to select a couple of key aspects of the performance to improve
22. The educator helped the learner to work out how they could improve their performance and specify the practical steps to achieve it
23. The educator checked if the learner understood their learning goals and action plan, by asking them to summarise it in their own words
24. The educator checked if the learner understood the rationale for their learning goals and action plan
25. The educator discussed with the learner possible subsequent opportunities for the learner to review their progress

3.6 Discussion

We sought to distinguish the distinct elements of an educator's role in feedback, endorsed by the literature, and to develop consensus on a set of observable behaviours that could engage, motivate and enable a learner to improve their performance in clinical practice. Support for these elements came from triangulating information from observational studies of feedback, surveys and interviews of educators and learners, summaries of written feedback forms, systematic reviews and meta-analyses of feedback, and established psychological and behavioural theories, in addition to expert argument, published across health professions, education, psychology and business literature. However there is little high quality evidence to substantiate these educator behaviours and they require formal testing to explore their impact in clinical practice. One of the drivers for this research was the desire to investigate whether specific constituents of feedback argued to be important, do indeed enhance learning.

Characteristics of educator feedback behaviours in high quality feedback

We identified 18 distinct elements and 25 educator behaviours; this exposes the complexity of a feedback interaction. To facilitate further discussion and consideration, we propose four overarching themes that may describe the key concepts of high quality feedback.

1. The learner has to 'do the learning'

A learner needs to develop a clear vision of the target performance, how it differs from their performance and the practical steps they can take to improve their subsequent performance (Statements: 14-16,22-24).[1,18,54] This requires the learner to make sense of an educator's comments, to compare the new information with their previous understanding of the issue and resolve gaps or discrepancies.[105,113,150] A learner has to actively construct their own understanding; an educator cannot deliver it 'ready-made' to them. Feedback is best done as soon as the learner and educator can engage after the performance (Statement: 2). A learner can only work on one or two changes at a time, in accordance with theories of cognitive load.[151] This would suggest that it is important to prioritise the most important and relevant issues (Statement: 21).[113,115] As feedback is an iterative process, the progress achieved (or difficulties encountered) after implementing the action plan should be reviewed (Statement: 25).[14,18,113]

The primary purpose of the learner's self-assessment is to develop their evaluative judgement, contributing to their self-regulatory skills (Statement:7-8,12-13).[152] The learner is positioned to take responsibility for their own learning. As they compare their performance to the target performance, it offers an opportunity for them to clarify their vision of the target performance (Statement: 14), calibrate their assessment to the educator's assessment (Statement: 15), and highlight their priorities and ideas about how their performance could be improved (Statement: 7-8).[14]

Once the learner is seen as 'the enacter' of feedback, the educator's role becomes 'the enabler'. The educator uses their expertise to discuss the performance gap, explore the learner's perspective and reasoning, clarify misunderstandings, help to solve problems, offer guidance in setting priorities and effective goals, and suggest ideas for improvement (multiple statements).

2. The learner is autonomous

High quality feedback supports a learner's intrinsic motivation to develop their expertise and respects their autonomy.[107] It recognises that the learner decides which changes to

make (if any) and how they will do this. Feedback information is only 'effective' if a learner chooses to implement it. This is more likely to occur when a learner believes an educator's comments are true and fair, and will help them to achieve their personal goals. This is more likely when an educator's comments are based on specific first-hand observations (Statement: 1) as a starting point for an open-minded discussion with the learner about the reasons for their actions, and enables identification of learning needs (Statements: 17-19).[3,153,154] An educator's comments are best directed to actions that can be changed, not personal characteristics (Statement: 20), that is, 'what the learner *did,* not what the learner *is*'.[2,3,154] Comments that target a person's sense of 'self' (including valued self-concepts like 'being a health professional') or general corrective comments, may stimulate strong defensive reactions, and do not appear to improve task performance.[2,35,155,156] To support a learner's intrinsic motivation, an educator should offer suggestions as opposed to giving directives, explain the reasons for their recommendations and help a learner to develop an action plan that aligns with their (often revised) goals, priorities and preferences (Statements: 7,14,18,22,24).[32,56,107]

3. The importance of the learner: educator relationship

The learner-educator relationship strongly influences face-to-face feedback; the personal interaction can enrich or diminish the potential for learning.[1,25,29] During the encounter, a learner's interpretation of the educator's message is affected by their knowledge and experience of the educator. If a learner believes an educator has the learner's 'best interests at heart' and is respectful and honest, this creates a trusting relationship and an environment that supports learning (Statements: 3-4,11).[29] This sense of trust, or psychological safety, encourages the learner to take a 'learning focus' not a 'performance focus', so the learner can concentrate on improving their skills, as opposed to trying to appear competent by covering up difficulties (Statement: 9).[113,155,157] Performance evaluation often stimulates emotions.[6] An educator may help by responding to a learner's emotions appropriately (Statement: 10).[158] In addition an educator should aim for a feedback process that is transparent and therefore predictable, which may help a learner manage feelings of anxiety about what is likely to happen in the session (Statements: 5).[124,159]

4. Collaboration

Collaboration, through dialogue, is essential for high quality feedback (multiple items). The learner and educator work together, with the common aim of creating an individually-tailored action plan to help the learner improve. The behaviours specified in the items are designed to promote shared understanding and decision-making. Feedback is more than two separate contributions; each one seeks, responds to and builds on the other's input. Face-to-face verbal feedback offers a unique opportunity for direct, immediate and flexible interaction. This makes it possible for a learner or educator to seek further information, clarify what was meant, raise different perspectives, debate the value of various options and modify proposals in response to the other's comments. Collaboration optimises the potential for a fruitful outcome because insufficient information, misunderstandings and other obstacles to success can be dealt with during the discussion.

Research strengths and limitations

This research has several strengths. It addresses an important gap in health professions education with a practice-orientated solution. The research design was systematic and rigorous. The literature search was extensive and continued to the point of saturation but we cannot be sure that all relevant information was assembled. Countering the potential for oversight was the in-depth scrutiny by experts in the health professions and education.

3.7 Conclusions

Work-based learning in the health professions relies on effective verbal feedback but problems with current feedback practice are common.[160] This research advances the feedback literature by creating an endorsed, explicit and comprehensive set of educator behaviours intended to engage, motivate and support a learner during a feedback interaction. The recommended educator behaviours provide a platform for developing a method to systematically evaluate the impact of the verbal feedback on learner performance.

3.8 Chapter 3 Closing discussion

The work reported in this chapter clarified informed decisions regarding 'what constitutes quality learner centred feedback', deepened understanding of the related theory and research underpinning those constituents and started the process of explaining how educators can promote quality face-to-face feedback. The task of identifying important components in effective feedback and developing descriptions of corresponding educator behaviours was complex and challenging. This may explain why this had not been accomplished previously.

The set of observable educator behaviours, described in the provisional instrument, was supported by a powerful combination of substantiating research information and expert consensus. However, it was likely to have shortcomings. For example, some items could be less important or less generally applicable, some may overlap with others, and the large number of items may make the instrument difficult to use in clinical practice. Although all items were endorsed during the final face-to-face meeting of the expert panel, the phrasing of a few items remained contested. This indicated the difficulty of unambiguously describing an observable behaviour that exemplified a sophisticated concept. One example was *Item 4: The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise.* The underlying concept targeted by this, called a 'growth mindset' by Dweck, accepts that mistakes and shortcomings are expected during learning and should be viewed as opportunities for further development.[157] Within the expert group, this was endorsed as an essential principle in effective feedback, yet the phrasing of this item remained awkward despite repeated revisions. As testing and further refinement would occur in Phase 3, all items ratified by the expert panel were retained in the next phase of the research.

At the time of finalising this thesis, this article had 33 citations (author citations excluded) demonstrating considerable interest in the findings.

In the next steps in this PhD, the provisional instrument was utilised to examine authentic face-to-face feedback discussions in clinical practice. This facilitated analysis of how easy the instrument was to administer and also provided data on the typical profiles of educators' current feedback practice in the health professions.

Chapter 4:
Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis

4.1 Chapter 4 Introduction

Phase 2 of this research, described in Chapter 3, resulted in the creation of a comprehensive suite of educator behaviours deemed to be important for enhancing learner outcomes. This set of items constituted a provisional feedback instrument, ready for testing in the next phase.

Phase 3, the final phase of this research, covered administration, analyses and refinement of the provisional instrument, to produce the Feedback Quality Instrument. The three major stages in this phase included 1) assembling a team of raters, each of whom administered the instrument to analyse videos of feedback interactions to generate ratings data and to conduct usability analysis (described in Chapter 4); 2) quantitative analysis, involving psychometric analysis of these rating data (described in Chapter 7), and parallel qualitative analysis of the video data; and 3) refinement of the instrument based on usability testing, psychometric analysis, parallel qualitative analysis and foundational research and theory (also described in Chapter 7). The parallel qualitative research explored two important phenomena, evaluative judgement and psychological safety, which were fairly new concepts within the health professions education literature and not well characterised in pre-existing literature (described in Chapters 5 and 6).

Phase 3, Stage 1, described in this chapter, focused on systematic analysis of face-to-face feedback discussions in routine clinical practice, using the provisional instrument. In order to capture authentic and varied feedback practice, the study design used video-recording of routine feedback episodes involving a professionally diverse range of

educator-learner pairs working together. This posed substantial challenges, particularly related to consenting legitimate educator-learner pairs, recording genuine feedback discussions taking place within busy clinical care environments and collecting the videos. The setting was a major metropolitan teaching hospital network incorporating six hospitals, so there was an abundance of potential participants. Recruitment started with the educators. To do this, widespread publicising of the study was conducted, evolving over time to include advertising posters, presenting at unit meetings and contacting senior clinicians via emails distributed by their unit administrative assistants, then following up with those who might be interested by telephone or in person. As the instrument focused on educators' practice, this placed the senior clinician in the unusual position of being 'the one under the spotlight'. Most educators expressed reluctance to record themselves during a feedback session and admitted they felt embarrassed at the thought of the ratings team, including the primary researcher (CEJ), observing their practice. The educators said they were unsure what they were supposed to be doing and lacked confidence in their feedback skills. The professional risk associated with the feedback videos, for both learners and educators, had been carefully considered during the study design. Hence various design features had been incorporated to limit this, such as requiring participants to delete their copy of the video once it had been successfully uploaded to a secure site, removing names from video transcripts and maintaining strict confidentiality within the research group. Yet the risk was perceived as still too high for some educators to participate.

Once an educator consented, the process was repeated to see whether one of the learners working with them was willing to participate, to complete recruitment of an educator-learner pair. This was conducted by the primary researcher (CEJ) not the educator, to minimise learners perceiving any pressure to participate. Then the pair needed to organise recording their next routine feedback session. It was decided to recommend that the pairs video themselves, as this would minimise difficulties related to scheduling someone else to record the session, particularly as feedback sessions are often fitted in whenever there is a brief lull in clinical work. This also avoided the intrusion of a third party. Educators frequently cited potential technical difficulties as

another major barrier to participation. While today recording and uploading videos using mobile phones is relatively commonplace, when this study started most health care professionals in the participating health service had little experience with recording themselves or uploading videos using any device. In preparation for the study, simple ways to video feedback discussions using smart phones or laptops, then upload them to a secure cloud-based storage, were devised, and clear instructions, with both written and visual information, were produced. Additional advice, provision of suitable equipment and in-person assistance was made freely available, to minimise frustration related to technical aspects of the study. Nevertheless, a variety of setbacks occurred which resulted in 5 failed recordings from a total of 41 (13%) due to the phone battery running out part way recording the interaction, accidental selection of time-lapse video format and unstable positioning of the recording device that led to participants not remaining in view.

Hence recruitment was protracted, taking eighteen months and requiring repeated refinements of enrolment strategies, and a revised application to the ethics committee, to collect a substantial number of videos. However, the final result was very rare observational data involving a diverse group of health professionals. There were a mix of men and women across all subgroups; educators with few or many years of experience as a clinical supervisor; learners ranging from students to specialist medical trainees and senior clinicians learning new skills in education or simulation-based education; and from different health professions including medicine (all major specialities), nursing and physiotherapy.

Once all the feedback videos were collected, the research team prepared to analyse the educators' practice in the 36 videos using the 25-item provisional instrument. In this way, usability testing, including consistency in application of items and rating of items, could be evaluated. In preparation for this, raters practised administering the instrument on videos, which were not part of the data set, using an initial 5 point item rating scale:

How well did the educator do this?

0= not adequate: not demonstrated to adequate standard.

1= borderline: demonstrated to an adequate standard in part.

2= adequate: demonstrated to an adequate standard.

3= good: demonstrated to a good standard.

4= outstanding: demonstrated to an excellent standard.

Even with discussion after these practice trials, raters found it very hard to discriminate each educator behaviour across the session to this degree. They agreed that it was only workable to rate whether the educator 'did not', 'partly did' or 'did' demonstrate the behaviour described in each item. Hence the rating scale was reduced accordingly to three options for the videos included in the analysis.

Despite these difficulties, this stage in the research resulted in a rich and informative body of data that was used to comprehensively summarise the components seen during typical feedback conversations in clinical practice and reported in the publication that follows.

Data analyses

The details regarding my contribution to the data analyses presented in the following chapter are detailed here. The descriptive statistics were performed by me, under the guidance of my supervisor Professor Jennifer Keating, using Microsoft Excel (Microsoft Office 365, Microsoft Corporation, Available at: https://office.microsoft.com/excel).

Chapter 4: Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis

This chapter is the study published in BMC Medical Education (2019) 19:129 (Lightly edited for headings and formatting consistency). Published May 2019.

Christina E. Johnson[1,2]; Jennifer L. Keating[3]; Melanie K. Farlie[4]; Fiona Kent[5]; Michelle Leech[6] and Elizabeth K. Molloy[2]

[1] Monash Doctors Education, Monash Health, Melbourne, Victoria, Australia.

[2] Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia.

[3] Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine Nursing and Health Science, Monash University, Melbourne, Australia.

[4] Allied Health Workforce Innovation, Strategy, Education & Research (WISER) Unit, Monash Health, and School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences at Monash University, Melbourne, Australia.

[5] Education Portfolio, Faculty Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia.

[6] Monash School of Medicine, Faculty of Medicine, Nursing & Health Sciences, Monash University and Monash Health, Melbourne, Australia.

Keywords: Feedback, Formative feedback, Effective feedback, Health professions education, Professional development

Acknowledgments: The authors are indebted to the educators and learners who so generously participated in this research.

Funding: Christina Johnson, Jennifer Keating and Elizabeth Molloy received the following grant: Office of Learning and Teaching (OLT) and Higher Education and Research Development Society in Australasia (HERSDA) Researching New Directions in Learning and Teaching: Seed Funding to Support Project Mentoring and Collaboration.

Declaration of interest: None

4.2 Abstract

Background: Verbal feedback plays a critical role in health professions education but it is not clear which components of effective feedback have been successfully translated from the literature into supervisory practice in the workplace, and which have not. The purpose of this study was to observe and systematically analyse educators' behaviours during authentic feedback episodes in contemporary clinical practice.

Methods: Educators and learners videoed themselves during formal feedback sessions in routine hospital training. Researchers compared educators' practice to a published set of 25 educator behaviours recommended for quality feedback. Individual educator behaviours were rated 0 = not seen, 1 = done somewhat, 2 = consistently done. To characterise individual educator's practice, their behaviour scores were summed. To describe how commonly each behaviour was observed across all the videos, mean scores were calculated.

Results: Researchers analysed 36 videos involving 34 educators (26 medical, 4 nursing, 4 physiotherapy professionals) and 35 learners across different health professions, specialties, levels of experience and gender. There was considerable variation in both educators' feedback practices, indicated by total scores for individual educators ranging from 5.7 to 34.2 (maximum possible 48), and how frequently specific feedback behaviours were seen across all the videos, indicated by mean scores for each behaviour ranging from 0.1 to 1.75 (maximum possible 2). Educators commonly provided performance analysis, described how the task should be performed, and were respectful and supportive. However a number of recommended feedback behaviours were rarely seen, such as clarifying the session purpose and expectations, promoting learner involvement, creating an action plan or arranging a subsequent review.

Conclusions: These findings clarify contemporary feedback practice and inform the design of educational initiatives to help health professional educators and learners to better realise the potential of feedback.

4.3 Background

Modern clinical training, aligned with competency based education and programmatic assessment, is focused on assessment and feedback on routine tasks in the workplace, targeting the highest level in Miller's framework for competency assessment.[46,47,161] Feedback is one of the most powerful influences on learning and performance.[1,2,33,34,162] It offers the opportunity for a learner to benefit from another practitioner's critique, reasoning, advice and support. Through this collaboration, the learner can enhance their understanding of what the performance targets are and how they can reach those standards.[6,29] 'On the run' or informal feedback refers to brief fragments of feedback that occur in the midst of delivering patient care. A formal feedback session typically refers to a senior clinician (educator) and student or junior clinician (learner) discussing the learner's performance in a more comprehensive fashion. Formal feedback sessions often occur as a mid- or end-of-attachment appraisal or as part of a workplace-based assessment. However the success of this model relies on everyday clinicians providing effective feedback. It is not clear which components of effective feedback have been successfully translated from the literature into supervisory practice in the workplace, and which have not. Information on gaps in translation could be used to better target professional development training, or to design strategies to overcome impediments to implementing quality feedback behaviours.

Studies involving direct observation of authentic feedback in hospitals are rare. Observational studies are highly valuable, as they provide primary evidence of what actually happens in everyday clinical education. Direct observation can be achieved either by researchers observing the activity or via video-observation. We identified only a few previous direct observation studies: these involved junior learners (medical students or junior residents) in a few specialties (internal or family medicine) involving formal or informal feedback (in outpatient clinics, on a ward, or following summative simulated clinical scenarios).[12,38,114,118,121,163-165] An additional single study involved physiotherapy students during formal mid- or end-of-attachment feedback.[115] The scarcity of observational studies may be related to the time consuming nature, difficulty in arranging observers or video recording to coincide with feedback meetings slotted

into busy schedules, or the reticence of participants to be observed or recorded. These studies reported that typically educators make comments on specific aspects of performance, teach important concepts, and describe or demonstrate how the learner can improve. However educators tend to speak most of the time, ask the learner for their self-assessment but then do not respond to it, avoid corrective comments and do not routinely create action plans. However these findings may no longer reflect current practice. In addition, no study captured the diversity of clinical educators and learners that work in a hospital environment.

Therefore we set out to directly observe authentic formal feedback episodes in hospital training, via self-recorded videos, to review contemporary educators' feedback practice in workplace-based learning environments. This could then clarify opportunities and inform the design of professional development training. In Australia, health professions training is concentrated in hospitals, integrating both inpatient wards and outpatient clinics; major dedicated specialist outpatient centres are rare and family medicine clinics are relatively small. We recruited a range of participants, characteristic of the diversity present in hospitals, as desirable feedback elements are not profession specific. We targeted formal feedback sessions to capture complete feedback interactions. We then analysed the composition of educators' feedback practice using a comprehensive set of observable educator behaviours recommended for high quality feedback (see Table 4.1).[52] This enabled a systematic analysis of the first set of data gathered using a comprehensive set of behavioural indicators, in contrast to previous studies in which less structured and more exploratory approaches were used. This framework outlines 25 discrete observable educator behaviours considered to enhance learner outcomes by engaging, motivating and assisting a learner to improve (see Table 4.1).[52] This earlier publication by our team described how these items were developed, starting with an extensive literature review to identify distinct elements of an educator's role substantiated by empirical information to enhance learner outcomes, then operationalised into observable behaviours and refined through a Delphi process with experts.

While we strongly endorse a learner-centred paradigm, we have chosen to focus on the educator's role in feedback because educators are in a position of influence to create conditions that encourage learners to feel safe, participate and work out how to successfully improve their skills. We agree that specific feedback episodes are shaped by the individuals involved, the context and the culture, however strategies to promote a learner's motivation and capability to enhance their performance remain relevant. Recommended feedback behaviours are not intended to be implemented in a robotic fashion but tailored to a particular situation by prioritising the most useful aspects throughout the interaction. The core segments of quality feedback include clarifying the target performance, analysing the learner's performance in comparison to this target, outlining practical steps to improve and planning how to review progress.[1,6,18] Overarching themes include promoting motivation,[32,54,56,107] active learning[105,150,152] and collaboration[113,166-168] within a safe learning environment.[3,29,55]

Research question

The research questions addressed in this study were:

1. What behaviours are exhibited by clinical educators in formal feedback sessions in hospital practice settings?

2. How closely do these behaviours align with published recommendations for feedback?

Table 4.1: Set of 25 educator behaviours that demonstrate high quality feedback in clinical practice.

Reproduced with permission from Johnson *et al*. *BMC Medical Education* (2016) 16:96 Identifying educator behaviours for high quality verbal feedback in health professions education: literature and expert refinement.

Orientation and Process *1. Based on observed performance* The educator's comments were based on observed performance *2. Timely feedback* The educator offered to discuss the performance as soon as practicable *3. Feedback purpose clear* The educator explained that the purpose of feedback is to help the learner improve their performance *4. Establish a non-judgmental atmosphere: 'here to help'* The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise *5. Clarify feedback process, so learner knows what to expect* The educator described the intended process for the feedback discussion
Learner-centred Focus *6. Encourage dialogue* The educator encouraged the learner to engage in interactive discussions *7. Seek learner's priorities* The educator asked the learner about their learning priorities for the observation and feedback discussion, and responded to them *8. Encourage learner to 'work it out for themselves'* The educator encouraged the learner to consider the issues and possible solutions during the feedback discussion *9. Encourage learner to focus on learning, rather than trying to cover up limitations* The educator encouraged the learner to discuss difficulties and ask questions regarding the performance so the educator could help the learner to develop solutions *10. Acknowledge learner's emotional response* The educator acknowledged and responded appropriately to emotions expressed by the learner *11. 'Best interests at heart'* The educator showed respect and support for the learner

Performance Analysis

12. Clarify the value of self-assessment
The educator asked what the learner understood about the benefits of self-assessment and helped clarify

13. Learner self-assessment
The educator asked the learner to identify key similarities and differences between the learner's performance and the target performance

14. Target performance and reasoning clear
The educator clarified with the learner key features of the target performance and explained the reasoning

15. Educator assessment, including clear performance gap
The educator clarified with the learner similarities and differences between the learner's performance and the target performance

16. Educator comments on a few, important issues
The educator's comments focused on key issues for improving the performance

17. Specific instance ('what happened')
First the educator described, using neutral language, what the learner did (action, decision or behaviour), and the consequences

18. Educator's perspective clear ('why it matters')
The educator clearly explained their perspective on the learner's actions, including the reason for their concern

19. Educator explores learner's perspective ('why' the learner acted as they did)
The educator explored the learner's perspective and reasoning to reveal the basis for the learner's actions (e.g. what was the learner trying to do and options considered/ difficulties encountered)

20. Focus on actions, not the person ('did' not 'is')
The educator's comments were focused on the learner's actions not personal characteristics

Action Plan

21. Select learning priorities: most useful (important and relevant) for the learner
The educator helped the learner to select a couple of key aspects of the performance to improve

22. Develop the action plan: how to do it!
The educator helped the learner to work how they could improve their performance and specify the practical steps to achieve it

23. Check the learner understands the plans
The educator checked if the learner understood their learning goals and action plan, by asking them to summarise it in their own words

24. Checks the learner understands the rationale: 'why it's better'
The educator checked if the learner understood the rationale for their learning goals and action plan

25. Plan opportunities to review the impact of the feedback
The educator discussed with the learner possible subsequent opportunities for the learner to review their progress

4.4 Methods

Research overview

In this observational study, senior clinicians (educators) observed junior clinicians or students (learners) performing routine clinical tasks in a hospital setting and then videoed themselves during the subsequent formal feedback session. We analysed each video using a check-list based on the set of educator behaviours recommended in high quality feedback (see Table 4.1).[52]

The feedback videos were captured at multiple hospitals within one of Australia's largest metropolitan teaching hospital networks between August 2015 and December 2016. Ethics approval was obtained from the health service (Reference 15233L) and university human research ethics committees (Reference 2015001338). Written informed consent was obtained by all participants.

Recruitment

Educators (senior clinicians) across medicine, nursing, physiotherapy, occupational therapy, speech therapy and social work, and their learners (either qualified health professionals undertaking further training or students), who were working with them, were invited to participate. A broad range of educators were sought, via widespread advertising of the study using flyers, emails circulated by unit administration assistants, short presentations at unit meetings and face-to-face meetings with staff across the health service. To be considered for participation, an educator had to contact the primary researcher (CEJ), in response to the advertisement. Once an educator consented, they distributed flyers to any learners working with them, with instructions to contact the primary researcher (CEJ) if the learners were interested in participating. Diversity was sought by rolling advertising to participants, with consideration of key factors including health profession and specialty, gender and supervisor experience (educators) or training level (learners). Once an educator and a learner had both consented, the pair were advised and they made arrangements to video a routine feedback session. They were asked to record an entire feedback encounter and aim for a duration of approximately ten minutes but were not given any additional instructions

regarding how to conduct the feedback session. Participants were not shown the set of 25 educator behaviours recommended for high quality feedback used to analyse the videos nor given any other education on feedback from the research team, as the aim was to study the nature of current feedback practices.

Consenting participants used a smart phone or computer to video-record themselves at their next scheduled formal feedback session related to either a workplace-based assessment or end-of-attachment performance appraisal. This video was subsequently uploaded to a password protected on-line drive and participants were instructed to delete their copy. The videos were numbered using a random number generator and the videos (other than the images) contained no personal identifying information.

Video analysis

The group of raters were all health professionals (two medical, four physiotherapy) in senior education/educational research roles with extensive experience in supervision and feedback. Each rater analysed each video independently and compared their observations with the set of 25 educator behaviours recommended for high quality feedback (see Table 4.1).[52] Each educator behaviour was rated 0 = not seen, 1 = done somewhat or done only sometimes, 2 = consistently done.

In a preparatory pilot study, we rated three videos using the instrument. We then met to discuss ratings and to identify differences in interpretation of items and the use of the rating scale. Strategies to encourage concordance and to clarify item meaning were developed. In particular we identified that Behaviour 2: *Timely feedback*: *The educator offered to discuss the performance as soon as practicable* was not observable, so it was excluded. For Behaviour 10: *Acknowledge learner's emotional response: The educator acknowledged and responded appropriately to emotions expressed by the learner,* we decided that this would be rated as '2' (consistently done) in the following situations i) if implicit or explicit indicators of learner emotion (such as anxiety or defensiveness) were detected, and the educator acknowledged, and attended to this, or ii) if emotional equilibrium was observed throughout the encounter, as we assumed that this emotional

balance between educator and learner required the educator to be reading cues and acting accordingly. Subsequently the total item score could range from 0 - 48.

Data analysis

The data provided two perspectives i) on an individual educator's practice: how many of the behaviours recommended in high quality feedback were observed in each video and ii) across the whole group of educators: which behaviours were commonly performed? To characterise each individual educator's practice seen in a video, the scores for each item were averaged across assessors and then summed to give a total score. To describe how commonly specific educator behaviours were observed amongst the whole group of educators, the mean score and standard devation for each item was calculated across all the videos.[169] To assess inter-rater reliability, total scores for each video were assessed for concordance between examiner pairs using Spearman's rho.

4.5 Results

Thirty-six feedback videos were available for analysis after five were excluded: two because they were incomplete (insufficient smartphone memory) and three because of technical errors with recording (audio unclear, time-lapse format used, participants not visible).

Video participants

Thirty-four educators participated, with diversity across key characteristics (health profession and specialty, length of supervisor experience and gender). There were four nurses, four physiotherapists and 26 senior medical staff (three anaesthetists, three emergency physicians, two radiologists, one paediatrician, six physicians, three psychiatrists, three obstetrician-gynaecologists, one opthalmologist and four surgeons). There were 18 (52.9%) female and 16 (47.1%) male educators. Fourteen (41.2%) educators had five years or less educator experience, 11 (32.3%) had six to ten years and 9 (26.5%) had more than ten years.

Thirty-five learners participated with diversity across key characteristics (health profession and specialty, training level and gender). There were 9 (25.7%) students, 9

(25.7%) clinicians who were five years or less post-qualification, 15 (42.9%) clinicians six years or more post-qualification and 2 (5.7%) senior clinicians. Twenty-three learners (65.7%) were female and 12 (34.3%) were male. All participants were from the same health profession and specialty as their respective educators.

The feedback session was related to a mid- or end-of-attachment assessment in 11 (30.6%) videos and to a specific task (such as a procedural skill, clinical assessment, case discussion or presentation) in 25 (69.4%) videos. An official feedback form from an institution such as a university or specialist medical college was used in 11 (30.6%) of the feedback sessions, most of which were mid- or end-of-attachment assessments. Most of the assessments were formative but some were summative as a component of longitudinal training programs aligned with programmatic assessment principles.[47]

Analysis of educator behaviours during feedback

Each video was analysed by four to six raters providing a total of 174 sets of ratings (unexpected time constraints on the project limited analysis by two raters). Missing data were uncommon (0.2% ratings missing).

Inter-rater reliability

To maximise data for comparison, the inter-rater reliability range for total scores was calculated for raters (4/6) who analysed all the videos: Spearman's rho was 0.62 - 0.73. The other two raters rated 10 (28%) and 21 (58%) of the 36 videos and were not included in the inter-rater reliability analysis.

Individual educator's feedback practice

To learn more about individual educator's practice and how many of the recommended educator behaviours were observed in each video, we calculated a total score (sum of rating for each observed behaviour, averaged across all assessors) for each video. Total scores ranged from a minimum of 5.7 (11.9%) to a maximum of 34.2 (71.3%), with a mean score across educators of 22.5 (46.9%, SD 6.6), from a maximum possible score of 48. More detailed analysis (see Table 4.2) revealed virtually all the educators (88%) had

a total score between 10 - 30. Although it was not our intention to compare performance across different characteristics (which would require sufficient sample sizes for each group, to enable comparisons), there seemed to be a fairly even spread of health professions, experience and gender across the score ranges.

Table 4.2: Range of total scores for individual educators (34 educators in 36 videos)

Total scores	Number of educators	Health Profession	Supervisor experience (years)	Gender
	Total=34	M: medicine N: nursing P: physiotherapy	≤ 5y | 6-10 | > 10y	F: female M: male
40.1-48	0	0	0	0
30.1-40.0	3 (9%)	2 M 1N	1 | 1 | 1	2F 1M
20.1-30.0	16* (50%)	12M 2N 2P	8 | 4 | 4	7F 9M
10.1-20.0	14* (38%)	11M 1N 2P	5 | 6 | 3	8F 6M
0-10.0	1 (3%)	1 M	0 | 0 | 1	1F

*included one educator who featured in two videos (mean total score used)

Frequency of specific educator behaviours across the whole group of educators

To explore how often specific feedback behaviours were observed amongst all participants, we calculated the mean rating score for each behaviour across all the videos. Table 4.3 displays the rating mean (SD) for each behaviour, ranked from most to least often observed. Some behaviours were seen in almost every video (highest mean rating 1.75, Behaviour 10) while others were very infrequently observed (lowest mean rating 0.05, Behaviour 25).

Amongst those educator behaviours most commonly observed (top third: mean rating score 1.41 - 2.0), most related to the educator's assessment of the learner's performance. Educators commonly linked comments regarding learner performance to the learner's actions (Behaviours 1, 17, 20), focused on important aspects for improvement (Behaviour 16), described similarities and differences between the

learner's performance and the target performance (Behaviour 15), and clarified what should be done and why (Behaviour 14). The other two behaviours commonly seen related to creating a safe learning environment. These included showing respect and support (Behaviour 11) and responding appropriately to emotions expressed by the learner (Behaviour 10).

The middle band of educator behaviours were seen intermittently (mean rating score 0.71 - 1.40) and related to educators encouraging learners to contribute their thoughts, opinions and ideas, and to reveal their uncertainties. These included encouraging the learner to participate in interactive discussions (Behaviour 6), try to work things out for themselves (Behaviour 8), analyse their own performance (Behaviour 13), reveal the reasoning behind their actions (Behaviour 19), raise difficulties and ask questions (Behaviour 9), and participate in choosing the most important aspects to improve (Behaviour 21) and practical ways to do this through an action plan (Behaviour 22).

The lowest band of educator behaviours were rarely seen (mean rating score 0 - 0.7) and primarily related to the set up and conclusion of a feedback session. At the start of the session, as part of creating a safe learning environment, the recommended educator behaviours included explicitly explaining that the purpose of the feedback was to help the learner improve (Behaviour 3), describing the proposed outline for the session (Behaviour 5), and stating their acceptance that mistakes are an inevitable part of the learning process (Behaviour 4). As part of the session conclusion or wrap-up, the recommended behaviours included checking a learner's understanding of the learning goals and action plan (Behaviours 23, 24), and discussing future opportunities to review progress, to promote ongoing learning (Behaviour 25). The other educator behaviours that were rarely seen included the educator incorporating the learner's learning priorities (Behaviour 7) and promoting the learner's understanding of the value of their self-assessment (Behaviour 12).

Table 4.3: Observed educator behaviours ranked in order of rating, with the highest at the top.

Educator Behaviour	Rating Mean (SD)
10. The educator acknowledged and responded appropriately to emotions expressed by the learner	1.75 (0.34)
20. The educator's comments were focused on the learner's actions, not personal characteristics	1.72 (0.35)
11. The educator showed respect and support for the learner	1.65 (0.28)
1. The educator's comments were based on observed performance	1.58 (0.33)
18. The educator clearly explained their perspective on the learner's actions, including the reason for their concern	1.52 (0.52)
17. First the educator described, using neutral language, what the learner did (action, decision or behaviour), and the consequences	1.50 (0.53)
14. The educator clarified with the learner key features of the target performance and explained the reasoning	1.46 (0.43)
16. The educator's comments focused on key issues for improving the performance	1.43 (0.52)
15. The educator clarified with the learner similarities and differences between the learner's performance and the target performance	1.41 (0.47)
6. The educator encouraged the learner to engage in interactive discussions	1.36 (0.64)
8. The educator encouraged the learner to consider the issues and possible solutions during the feedback discussion	1.06 (0.64)
21. The educator helped the learner to select a couple of key aspects of the performance to improve	1.02 (0.55)
13. The educator asked the learner to identify key similarities and differences between the learner's performance and the target performance	0.97 (0.54)
9. The educator encouraged the learner to discuss difficulties and ask questions regarding the performance so the educator could help the learner to develop solutions	0.94 (0.58)
22. The educator helped the learner to work how they could improve their performance and specify the practical steps to achieve it	0.88 (0.49)
19. The educator explored the learner's perspective and reasoning to reveal the basis for the learner's actions	0.74 (0.63)
12. The educator asked what the learner understood about the benefits of self-assessment and helped clarify	0.35 (0.32)
7. The educator asked the learner about their learning priorities for the observation and feedback discussion, and responded to them	0.32 (0.27)

Educator Behaviour	Rating Mean (SD)
23. The educator checked if the learner understood their learning goals and action plan, by asking them to summarise it in their own words	0.29 (0.44)
4. The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise	0.26 (0.28)
5. The educator described the intended process for the feedback discussion	0.16 (0.28)
24. The educator checked if the learner understood the rationale for their learning goals and action plan	0.16 (0.29)
3. The educator explained that the purpose of feedback is to help the learner improve their performance	0.09 (0.16)
25. The educator discussed with the learner possible subsequent opportunities for the learner to review their progress	0.05 (0.11)
Total (standard deviation) Maximum score 48	22.5 (6.6) 46.9%
Total excluding Behaviours 17-19 Maximum score 42	18.9 (5.5) 45.0%

Rating scale:
0 = not seen, 1 = done somewhat or done only sometimes, 2 = consistently done

Colour coding:
Commonly seen (mean rating 1.41-2)
Sometimes seen (mean rating 0.71-1.40)
Rarely seen (mean rating 0-0.70)

4.6 Discussion

In this study of educators' feedback practice, we found considerable variation in both an individual educator's practice and how frequently specific recommended behaviours were observed across the group of educators. This provides valuable insights into 'what currently happens' during formal feedback episodes in hospital-based training. These insights clarify opportunities for future research into educator development with the potential for substantial impact. Furthermore the recommended behaviours offer a repertoire of specific strategies that may assist educators to understand and enact these quality standards.

Frequency of specific recommended behaviours observed across the group of educators

We found that educators routinely gave their assessment of the learner's performance and described what the task should look like, but only intermittently asked learners for self-assessment or development of an action plan. This seems to reflect a culture in which the educator's analysis of the learner's performance predominates.[23] These findings echo those from earlier observational studies and feedback forms.[38,114,115,117,119,163,170,171] This suggests that typical feedback practice in the clinical setting has remained much the same since these omissions were last reported years ago.

Self-assessment is a key component in self-regulated learning and evaluative judgement, which promotes reflection, independent learning and achievement.[113,152,166] Invitations for learner self-asssessment provide learners with the opportunity to judge their work first and indicate what they most want help with.[3,154,159] Self-assessments can alert the educator to the potential for a negative emotional reaction and rejection of the educator's opinion if the learner rates their performance much higher than the educator.[37] Self -assessment offers opportunities for learners to enhance their evaluative judgement by calibrating their understanding against an expert's understanding of the observed performance and the desired performance standards.[1,172] Recent work on student feedback literacy has highlighted the importance of strategically designing opportunities for learners to make judgements and discuss characteristics of quality work, to assist them to appreciate, interpret and utilise feedback.[26]

The fact that an action plan continues to be frequently neglected similarly warrants serious attention. If educators do not support and guide learners to create an action plan, learners are left with the difficult task of working out by themselves how to transform feedback information into performance improvement.[18] Furthermore, when learners hear about performance gaps, their distress may be exacerbated if they do not know how to improve it.[35]

Our study also identified a number of missing feedback features, which have not been previously documented. One involves positioning the development of a learner's motivation, understanding and skills as the focal point for feedback. The literature suggests that a learner is only likely to successfully implement changes when they 'wish to' (motivation) and 'know how to' (clear understanding).[6,11,113,173]

Self-determination theory argues that intrinsic motivation, which is associated with both higher performance and increased well-being, is promoted when a learner decides what to do, in line with their personal values and aspirations.[32,56,107] This is captured by recommended educator behaviours that position the learner as decision maker and the educator as guide (see Table 4.1: Behaviours 7, 21, 22). A learner must be convinced for themselves that the feedback is credible and valuable (Behaviours 1, 6, 7, 9, 20, 24).[34,99,100] The free flow of information, opinion and ideas between the educator and learner creates a shared understanding, as a foundation for tailored advice and good decision making.[174] In addition, Goal Setting Theory asserts that a learner's motivation is stimulated by a clear view of the performance gap, performance goals that are specific, achievable and valuable to the learner, and an action plan that is practical and tailored to suit their needs (Behaviours 14, 15, 21, 22).[54]

Recent advances in feedback have focused on the need to assist learners to process and utilise feedback information, so they 'know how to' enhance their performance. This is exemplified in the R2C2 feedback model, which includes assisting a learner to explore the information, their reactions to it and to design effective strategies for skill development.[166,168,174] Social constructivist learning theory describes how a learner makes meaning of new information through interactions with others.[20] To promote this active learning, recommended educator behaviours include encouraging the learner to analyse their own performance and 'work things out for themselves' (Behaviours 8, 12, 13), enquiring about the learner's difficulties or questions (Behaviour 9) and checking the learner's understanding of the action plan before concluding the session (Behaviours 23, 24).[98]

Another feature of effective feedback rarely seen in our study was educators deliberately setting up a safe learning environment at the start the session, although they showed respect and support for learners in general. Recent literature has reinforced the importance of promoting a safe learning environment and establishing an educational alliance.[55] This may be a particularly important strategy when the educator and learner do not have an established relationship, which seems to be increasingly commonplace in modern workplace training with short placements and multiple supervisors attending to learners.[175] Excessive anxiety negatively impacts on thinking, learning and memory.[98,176,177] Feedback is inherently psychologically risky; if a learner's limitations are exposed, this can result in a lower grade or critical remarks from the educator, or threaten a learner's sense of self.[2,3,35] Carless[29] highlighted the important role of trust in view of the strong relational, emotional and motivational influences of feedback. In an attempt to counter the natural anxiety, educators could be explicit that "mistakes are part of the skill-acquisition process" and that they desire to help, not to be critical.[98] In addition, if an educator negotiated the process and expectations for the feedback session, this could reduce the anxiety caused when the learner does not know, or have any control over, what is going to happen.[166]

One final important feature was the isolation of the learning activity. In our study, no educator discussed when or how the learner might be able to review to what extent they had been able to successfully develop the targeted skills (Behaviour 25); this was the lowest ranked behaviour of all. Molloy and Boud[6] have emphasised the importance of promoting performance development by linking learning activities, so that feedback plans can be implemented and progress evaluated in subsequent tasks. As supervision is increasingly short-term and fragmented in nature, collaborating with the learner in deliberately planning another opportunity to be assessed performing a similar task seems an important objective.

Individual educator's practice

The range in individual educator's scores found in our study suggests the educators had variable expertise in feedback. Educators were not shown the check-list of

recommended behaviours used in video analysis. Although not formally tested, there was no indication in the data that more experience conferred greater expertise, based on the spread of supervisor experience across the score ranges (Table 4.2). We did not ask about our educators' professional development training. Although potentially interesting, this information was tangential to our primary goal of assessing current workplace practice against recommended behaviours. Given that education paradigms have changed considerably across time, and that educator behaviour may partly reflect methods used when they were learners, the observed variability in feedback approaches highlights the need for continuing professional development that focuses on recent advances. The lack of striking differences in scores between professions suggests that feedback skills within formal encounters may be more similar than different. Hence feedback literacy training could, at least in part, be designed for educators across the health professions, allowing significant efficiencies. Nevertheless, the extent to which these skills vary within informal feedback encounters and across different contexts requires more study. Practising clinicians are responsible for the majority of health professions training (both senior students and junior clinicians) and yet specified standards for their education and training role are rare. In contrast health professionals spend many years training and being carefully assessed on their clinical skills.

The aim of our research is to assist educators in generating high quality learner-centred feedback, by developing descriptions of educator behaviours that could engage, motivate and enable learners to improve. It may well be that once clinicians have the opportunity to consider the recommended behaviours, it would be relatively easy for them to introduce missing elements into their practice. One strategy that might be valuable for educators would be to video their feedback with a learner and subsequently use the list to systematically analyse their own behaviours. This would enable educators to also engage in reflective learning and goal setting.[178,179] In addition, exemplars of supervisors' phrases or videos re-enacting quality feedback practices may help educators to translate the principles of high quality feedback into new rituals. The set of behaviours is comprehensive however it could be useful to prioritise or summarise

them, as twenty five recommended behaviours may seem overwhelming, especially to new educators.

Study strengths and limitations

Strengths of our study include self-recorded video-observations of authentic feedback episodes in routine clinical practice, to reveal 'what actually happens' and target the top level of Miller's framework for competency assessment. Participants involved a diverse group of clinical educators, characteristic of hospital practice. The educators' feedback practices were systematically analysed utilising an empirically derived, comprehensive set of 25 observable educator behaviours.

There are a number of limitations to our study. The small sample of 36 participants were from a single health service, although it is one of the largest in Australia with multiple hospitals. Participants volunteered (which may have resulted in a subset of educators and learners with stronger skills than those who did not volunteer) and participants recorded their own performances, potentially making our data overly optimistic. These factors limit the generalisability of our findings. In the application of the educator behaviour descriptions to the assessment of educator behaviour during feedback, there was some variation in rater consistency. One reason for this could be different interpretations of the educator behaviour descriptions. In future research, attention will be directed to refining the descriptions of observable behaviours and supporting information, accompanied by additional practice and discussion to optimise consensus amongst raters. Although video raters represented only two health professions (two physicians and four physiotherapists), which could raise the possibility that this might influence their analysis of educators' behaviours beyond their own profession, we cannot see a plausible argument to support this. A number of educators used official feedback forms (from an university, hospital or specialty college). Trying to complete these forms in accordance with their instructions, may have influenced educators' conduct or may have distracted educators' attention, as they can be quite cognitively demanding. However, there are no compelling reasons why best practice in feedback could not occur in parallel with any learner assessment rubric. In addition, educator-

learner pairs could have had earlier feedback conversations, during which some of the quality feedback behaviours may have occurred, particularly relating to setting up expectations and establishing trust, but were not captured on video.

4.7 Conclusions

Our study showed that during formal feedback sessions, educators routinely provided their analysis of the learner's performance, described how the task should be performed, and were respectful and supportive within the conversation. These are all valuable and recommended components of quality feedback. Nevertheless, other desirable behaviours were rarely observed. Important elements that were often omitted included deliberately instigating a safe learning environment at the start of the feedback session (by explicitly articulating the purpose, expectations and likely structure of the session), encouraging self-assessment, activating the learner's motivation and understanding, creating an action plan and planning a subsequent performance review. This suggests that many advances in feedback research, regarding the importance of assisting learners to understand, incorporate and act on performance information, have not impacted routine clinical education. Our research clarifies valuable targets for educator feedback skill development across the health professions education community. However further research is required to investigate whether implementing these recommended educator behaviours results in enhanced learner outcomes, as designed.

4.8 Chapter 4 Closing Discussion

Phase 3, Stage 1 of this research collected rare observational data of authentic feedback interactions in clinical contexts. Quantitative analysis revealed that individual educator's feedback practice varied substantially in the extent to which they demonstrated behaviours recommended in the instrument. The lack of a consistent standard of practice reinforced the potential value of clearly specifying useful behaviours. Presenting this data may motivate clinicians, hospitals and educational institutions regarding the value of need for professional development initiatives promoting feedback literacy and feedback skills, as the patterns of interactions deviated significantly from principles of good practice. The analysis also identified which educator behaviours, recommended in quality feedback, were less commonly seen in contemporary clinical practice. Some quality feedback behaviours were very common, such as performance analysis and clarifying features of the target performance but others were rare, such as seeking to find out a learner's priorities for the feedback discussion or planning opportunities to review subsequent progress. This appears to suggest that recommendations around 'educators telling learners how they did and what they should do' have been adopted but not those around 'involving the learner' and 'looking for the effect of feedback and linking learning episodes'. In particular, if educators assisted learners to review the success of action plans developed previously and then refine them for subsequent use, this could promote continuity of learning by strengthening the connection between discrete episodes of learning, commonly involving different educators. Gaining a better understanding of current patterns of feedback practice could inform priorities for educational interventions designed to boost feedback literacy for educators, and potentially for learners too, by addressing those aspects that most require improvement. Empirically-informed illustrative exemplars could be developed to show clinicians 'how to do this well', including the ways in which learners may meet, or deflect the 'invitations' from educators for their perspective.

Administering the provisional instrument to analyse educator behaviours seen in the feedback videos generated data that served a number of functions. As the raters

administered the instrument, they participated in usability analysis. The ratings data was used in two ways. First it formed the basis for the observational report, described in this chapter, that characterised and summarised contemporary educators' feedback practice. Secondly, it enabled subsequent psychometric analysis of the provisional instrument, as described in later chapters.

The videos provided a rich source of data that could be analysed in many ways. Phase 3, Stage 2 involved both qualitative and further quantitative analyses aimed at enhancing the provisional instrument. Two qualitative analyses on topical concepts are reported in this thesis. The first involved a study of how feedback encounters could serve to build learner evaluative judgment, which is reported in the chapter that follows, Chapter 5. The second concerned psychological safety of the learner during feedback exchanges, which is reported in Chapter 6.

Chapter 5:
Building evaluative judgement through the process of feedback

5.1 Chapter 5 Introduction

The preceding chapters outlined Phase 1, clarifying the current evidence on the impact of face-to-face feedback involving health professionals on clinical task performance (Chapter 2); Phase 2, developing a provisional feedback instrument (Chapter 3); and Phase 3, Stage 1, using the provisional instrument to analyse educators' practice seen in videos of authentic feedback interactions (Chapter 4). In the following chapter, Phase 3, Stage 2 is described, which focused on the process of investigating how to improve the provisional instrument, using qualitative and additional quantitative analyses. The following two chapters (Chapter 5 and 6) present qualitative analysis of the feedback video transcripts, to explore two important and emerging, yet under-researched areas related to promoting learner-centred feedback. The thematic analyses investigated how educators collaborated with learners to promote 1) learners' skills in evaluative judgement, in this chapter and 2) a psychologically safe learning environment for feedback, in the next chapter. These thematic analyses enabled the refinement of the provisional instrument by clarifying empirically informed, practical approaches for educators to support learner-centred quality feedback interactions.

This particular study, the focus of Chapter 5, was published in a book entitled 'Developing Evaluative Judgement in Higher Education'. The book was edited by colleagues at Deakin University who were leaders in conceptualising the construct of evaluative judgement. Evaluative judgement can be defined as 'understanding what constitutes high-quality work and applying these standards'.[180] The book explains the history and evolution of the construct, key ideas concerning evaluative judgement and how common practices could be used to develop this capability. Sadler first championed the expectation that feedback, to be effective, must enable learners to have a clear concept of the standards and how their own work compares.[18] All health professionals

require a commitment to high standards of practice and life-long learning so this skill, to appreciate 'what good work looks like' for themselves and others, is essential. Even when there are explicit standards, applying these to the disparate contexts across healthcare requires health professionals to appreciate complex implicit standards. The invitation to contribute to this book provided a timely opportunity to report on these qualitative findings, honing in on the prompts educators use to solicit learners' own evaluative judgement in feedback sessions, and the extent to which learners take up these invitations to make sense of their own performance.

Data analysis

The details regarding my contribution to the data analysis presented in the following chapter are detailed here. The video data was transcribed verbatim and was interrogated using thematic analysis. I independently read and open-coded a sub-set of transcripts in parallel with members of the core research team (EKM and JLK), and once our coding framework was collectively established, I led the thematic analysis of subsequent transcripts and utilised NVivo software (NVivo Windows 12.5.0, QSR International).

Building evaluative judgement through the process of feedback

This chapter is the book chapter published in *Developing evaluative judgement in higher education: assessment for knowing and producing quality work.* Edited by David Boud, Rola Ajjawi, Phillip Dawson, Joanna Tai. Routledge London. (Lightly edited for headings and formatting consistency). Published in April 2018.

Christina Johnson[1,2]; Elizabeth Molloy [2]

[1]Monash Doctors Education, Monash Health, Melbourne, Victoria, Australia.

[2]Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia.

Acknowledgements: We would like to thank the health professionals who volunteered to record their feedback practice. In the name of improvement, these participants rendered themselves vulnerable in allowing their behaviours to be scrutinised and evaluated: precisely the practice we argue for here.

5.2 Background

This chapter focuses on the potential for learners to develop evaluative judgement during feedback conversations in healthcare, as an illustrative example for any discipline. Feedback in the workplace and the classroom is often enacted as a once off provision of information to a learner, with minimal engagement of the learner in judgement-making or goal-setting.[115,181] We draw on the feedback literature and an empirical case study to argue that feedback can be an important mechanism in cultivating evaluative judgement skills in learners, when it is designed to develop those skills as well as improving performance, across sequential tasks.

Workplace learning is ubiquitous in healthcare and feedback is commonly held as one of the most important processes in enabling individuals to improve their practice.[3,5,50] Students learn alongside qualified clinicians amid the hustle and bustle of hospitals and community-based facilities[182] and following qualification, many health professionals undertake further training to gain specialist credentials.

Like all professionals in the workplace, health professionals need to develop self-regulatory skills that enable them to learn and perform independently as supervisors are not constantly available, especially post-qualification. These skills in self-regulation include how to identify learning goals, select effective learning strategies, monitor progress and then refine strategies accordingly. Evaluative judgment, understanding what constitutes high quality work and applying these standards, is an essential component of this.[180]

Feedback in Theory

Boud and Molloy defined feedback as 'a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work, in order to generate improved work'.[19] This definition aligns with a constructivist view of learning by positioning the learner as an active seeker of information, which they use to improve

their work. It aligns with the standpoint that feedback effectiveness can only be determined by its impact on subsequent performance.

The aim of feedback is to help the learner enhance their understanding and skills.[3] Effective feedback clarifies what the learner is aiming for (the characteristics of a quality performance), how their current performance compares to this standard and what they can do to improve.[1] Understanding the characteristics of quality work is critical to the development of evaluative judgement.[183] Standards, protocols guidelines, and rubrics can help provide learners with more information to clarify what constitutes 'good work'. However additional tacit knowledge is required to enable learners to distinguish what good work looks like in a particular context.[184] Molloy's[115]observational study of feedback in the workplace demonstrated that an important role for educators is to 'unpack' and 'bring to life' standards of work during feedback conversations.

Feedback in practice

Feedback sessions play an important role in the healthcare workplace. Typically an educator (senior clinician) meets with a learner (student, qualified junior or peer clinician) to review their performance on clinical tasks. This may relate to a specific performance, such as the learner's assessment of an unwell patient, or a longitudinal assessment of the learner's performance. Observational studies have shown that feedback discussions are typically dominated by the educator, and are rarely followed up upon.[114,115,118] This monologic feedback style positions the learner in a passive role and assumes that learners automatically enact desired changes. However, studies of feedback in the workplace and in the classroom setting, indicate that learners often do not understand feedback comments, or struggle to apply strategies to improve their work.[1,19] In addition, studies report that educators are reluctant to raise performance problems as they fear the learner may feel insulted or demoralised, with the potential to damage their collegial relationship.[12,30,115] These characteristics of one-way feedback, once-off feedback, and anxiety from both parties mean that feedback in practice often deviates considerably from the definition of feedback we ascribe to in this chapter.

An emerging literature suggests that verbal feedback can assist learners to develop and refine their own judgement.[14,115,180] Self-assessment is frequently inaccurate with high performing learners often underestimating their performance and low performing learners overestimating their performance.[185] The capacity for making accurate judgements can be improved by exposure to peers' performance (for the most competent, who underestimate) or enhancing their performance skills (for the least competent, who overestimate).[183,185] One of the key aspects to improving judgement is the opportunity to evaluate work and to observe how this self-generated judgement compares to the judgements of others (e.g. educator, peer, or consumer). If both learners and educators view feedback as a process with potential to improve learners' evaluative judgement (a broader goal than improving performance on the next task), we may begin to see different patterns of feedback enactment. The case study below identifies features of feedback that may influence learners' capacity to cultivate evaluative habits.

5.3 Building evaluative judgement during feedback: A case study in healthcare

Our research team aimed to find out more about how learners' evaluative judgement could be enhanced as part of verbal feedback processes in the workplace. We previously developed a feedback instrument through a literature review and Delphi process, which describes educator behaviours that could engage, motivate and enable a learner to improve their work.[52] Many of the items relate to the notion of evaluative judgement and encourage the learner to consider how their performance compares with a quality performance (the standard) and what they can do to improve subsequent performance.

As part of testing and refining the feedback instrument, we collected thirty six videos of authentic verbal feedback episodes involving health professionals during routine clinical practice. In order to examine opportunities for learners to practise and develop evaluative judgement, we performed an indepth analysis of ten of these videos, selected to maximise diversity across health professions, level of experience and gender. Using thematic analysis, we investigated i) what opportunities learners had to judge their own

work and how they engaged with those opportunities, and ii) what facilitated or impeded learners' development of evaluative judgement during feedback. Our three major findings are presented below.

1: Educator's invitations: I want to know what *you think* about the quality of your work

First we looked at how educators created opportunities for learners to evaluate their work. This was a very characteristic opening exchange:

> *Educator: Why don't you start off by telling me how you went?*
>
> *Learner: I thought it was OK…nothing went terribly wrong" [V11]*

Most educators initiated the feedback session by asking learners for their overall perspective. Learners almost universally responded by simply stating that their performance was adequate. Only once did a learner offer some genuine self-analysis by specifying features they had performed successfully at this point.

Following this initial invitation, some educators asked further questions, inviting the learner to evaluate their performance more deeply. This was a common approach:

> *Educator: What did you do well?..... If you had your time again, is there anything you would have done differently? [V9]*

Some educators employed other techniques to prompt learners to discuss their performance. One educator described a problem they had observed during the performance and then asked the learner for their perspective on what had happened:

> *On initial insertion [of the breathing tube during anaesthesia], there was a couple of coughs. What do you think was the issue there? [V11]*

Another educator asked the learner to describe how they had introduced themselves to the family, which prompted the learner to analyse their performance as they replayed

their steps. In another instance, the educator asked the learner to tell them about the standard for hand cleansing; as the learner described what should occur, they compared their own performance to it, highlighting the missed steps.

In contrast to these educators who probed further, some educators moved on to outlining their own assessment of the learner's performance straight after the opening exchange. This suggests educators did not see value in soliciting learners' self assessments or perspectives. As an extreme example, in one video the educator did not offer any cues for the learner to contribute and the learner did not speak until right at the end of the session, when they simply said, 'Thank you'.

Overall, educators created limited opportunities for learners to make evaluative judgements. Despite most educators providing an opening invitation to learners for self analysis, learners' responses were brief, superficial and deflecting. They did not offer a genuine analysis of their own performance and their responses echoed the generic nature of the educators' enquiries. In particular, educators did not explicitly ask learners to compare their performance to a standard or target performance. This could prompt the learner to undertake a more detailed analysis by specifying the aspects of their performance that met the standard and those that did not, as well as revealing the learner's understanding of the expected standard.

Learners appeared reluctant to offer self-evaluations. It is important to acknowledge that learners may well be making internalised judgements that they do not express, possibly because they are concerned about articulating flawed judgements or highlighting performance deficits undetected by the educator, or contravening perceived cultural norms where 'hearing it from the expert' is a shared expectation from both parties.[6,111,186]

2: Good work looks like this

The second finding explores how the characteristics of quality work were elucidated during the feedback session. In particular we consider how educators assisted learners to understand both the explicit and implicit criteria for quality work applied in a specific

context. In our sample, educators focussed on describing how the task should be performed. Educators' explanations about 'good clinical practice' were characteristically specific and straightforward. They often explained the reasoning that supported their recommendation; adding the 'why' to the 'what'. Regarding an emergency caesarean section operation for a woman in labour, an educator said:

> *My advice to you is to always make an adequate incision, that way you won't get stuck not being able to deliver the baby…and for labouring caesareans, you can always cut higher than you think. [V12]*

Regarding a clinical case presentation, an educator started by contrasting how expectations differ with seniority and setting:

> *The presentation of a case to a senior when you're a medical student is very different to when you're a training registrar…is again very different when you're a consultant presenting to a meeting…In the clinical setting, where there often isn't a lot of time…succinctness becomes really important. [V3]*

To help learners make sense of complex recommendations, educators often used exemplars of their own practice. During a discussion about how a junior clinician could engage with a confused elderly patient and stop the family interrupting without dismissing their important contributions, an educator gave examples of what they might say themselves in that situation:

> *I'll often say to the relative, 'I understand that you may have things that you know, that your relative doesn't know…and I'm very interested…. I'm going to talk to [the patient] first and then after that, you'll have the opportunity to tell me things that you think I need to know'. [V9]*

Although educators' discourse often included descriptions of what 'good performance' might look like, it was rare for learners to take the inititiative and ask educators for ways

in which they could improve their own performance, and even rarer for learners to suggest their own strategies.

3: Calibration of the learner's judgement through dialogue

Verbal feedback offers a valuable opportunity for educators to assist learners to calibrate their own judgements. This calibration can be achieved by educators confirming or contesting learners' evaluative comments. In one example, the educator enthusiastically validated the learner's opinion:

> *Educator: What do you think is an area…for development?*
>
> *Learner: I think being able to be more confident in my practice.*
>
> *Educator: Good! You must have read my notes; previously I said, 'Develop confidence in your own ability.' [V36]*

However educators seldom challenged learners' views, even though this is just as important for calibration as endorsing accurate judgements. An educator's performance analysis also enables a learner to compare their opinions (whether expressed or withheld) with an expert's opinion. In our case study, educators consistently made clear and specific comments regarding effective aspects of learner performance:

> *Educator: You were very engaging in your talk and your body language was good and you made eye contact with everyone. [V34]*

In contrast, educators' corrective comments were characterised by hints, timidity and interspersed by 'uhms', signifying the difficulties encountered in articulating their thoughts. When pointing out deficits, their body language suggested discomfort, such as leaning back from the table, fidgeting and avoiding eye contact.

> *Educator: **Uhm, in terms of the other part** of it, **so the other part** is really, **uhm**, was around the, **uhm, the, the** diagnosis, **uhm**, and again **some of the - some of the, uhm**, issues in relation to, **uhm, the, the** way you give feedback to the patient*

and, and *the understanding - **perhaps** a more direct feedback about whether there's been adequate medication trials, **uhm**, as to one of the reasons why **the, the** treatment hasn't worked. [V6]*

Our case study suggested that some educators assisted in calibrating learners' evaluative judgements by endorsing or (rarely) contesting learners' comments, and expressing their own evaluations. Educators frequently commended learners' quality performance or accurate judgements in a straightforward fashion. However educators tended to avoid corrective comments, and when they did attempt to discuss substandard aspects of performance, the language and non-verbal signals changed markedly to become apprehensive, apologetic, and circuitous.

Implications for practice

Evaluative judgement is a valuable skill for all those involved in life-long learning. If verbal feedback sessions are designed to engage learners, they can potentially act as prime occasions for educators to assist learners in developing their judgement capabilities.

In our case study, no learner spontaneously offered a thorough self-assessment and, even when asked, learners matched the depth of their evaluation to that of the educator's enquiry. In addition educators rarely paid much attention to learners' evaluations. This suggests neither learners nor educators saw the value of self-assessment, as a demonstration of learners' evaluation skills and understanding of 'good work'. As highlighted earlier, learners may be trying to avoid revealing their internal judgements, in case they are inaccurate or highlight undetected performance errors. Another likely reason for this pattern, is that learners may not realise that their superficial self-evaluation is insufficient (reinforced by educators immediately plunging into a monologue). Therefore it is likely that educators need to be better skilled in coaxing learners to share their authentic opinions. In the case of undergraduate and postgraduate students, assessment can be structured so that evaluative judgements are required from learners, and those attempts and explicit efforts to incorporate new strategies into subsequent tasks are rewarded.

One important factor influencing learner self-disclosure may be the level of trust in the learner-teacher relationship.[6,11,29,35,99,187,188] highlights the importance of 'the creation of a supportive, safe and trusting environment in which student learning is facilitated and risks can be taken'. Interestingly, in our selected sample of videos, no educators showed any vulnerability in expressing their own challenges or learning deficits at a similar stage of training or in current practice. This reciprocal vulnerability in face-to-face feedback has been hypothesised as an important ingredient in inviting honest exchange of information.[30]

Further studies that explore participants' intentions and motivations during and after feedback sessions would be valuable in helping design strategies to embed learners' articulation of genuine self-assessments into routine practice.

So often, standards of work in the classroom and the workplace are implicit[184,189]), so learners are more reliant on educators to assist them to gain an understanding of what they should be aiming for. In our case study, clarifying 'what good work looks like' was a major focus for educators in feedback. However, educators rarely sought questions from the learner or checked their understanding of concepts, implying educators believed they could easily transmit complex ideas. This finding aligns with Boud and Molloy's[19] description of current feedback practices which resembles educators delivering 'hopefully useful information'. A constructivist view on learning sees a learner making sense of new information and incorporating it into their mental model, by relating new information to their current understanding, and resolving any discrepancies.[113] There are multiple barriers to learners' accurate understanding of how they performed a task and how to better tackle a task. The educators' remarks may not be clear, specific or actionable, and even if they are, learners often need to ask questions or obtain further information to achieve an understanding of how to apply new ideas in practice. If educators viewed feedback as 'more enabling and less telling' this could change the focus from what educators say, to what learners understand.[6,52,190]

An earlier observational study of verbal feedback sessions in the hospital workplace described how educators socialise learners into the rituals of the professional

community.[115] Feedback appeared to be a one-way transmission of information from educator to learner. The consistency of this pattern suggests that this is an expected ritual between learners and educators. Calibration of learners' evaluation of their own and others' work is a critical element in the development of evaluative judgement. However in our case study, learners' evaluations were frequently not sought in depth or discussed, and so opportunities to calibrate learners' judgements by confirming or contesting them were often missed.

Educators typically find praise or validation of student performance more comfortable than highlighting aspects for improvement. It may be that educators highlight success in order to recognise learners' hard work and progress, as a key motivator for them to continue their endeavours.[2,54,56] Studies have also suggested that educators are concerned about the potential damage to relationships if the dialogue takes a critical tone.[115,191] In our case study, educators adopted a hesitant and apologetic tone when correcting learners' performance, as they 'tip-toed' around the topic. This reluctance to broach contentious topics reflects feedback in observational studies previously described as 'vanishing'[3] and 'mealy mouthed' feedback.[30] We need to consider how these patterns of interaction impact on a learner's development of evaluative judgement: with so much 'code' and 'padding', how can learners' decipher important messages that would help to shape their own judgement?

Based on research on learner-centred feedback practices and the empirical case study presented in this chapter, we make the following recommendations to improve evaluative judgement opportunities through verbal feedback:

1. The development of frameworks that describe desirable behaviours for both learners and educators in feedback: to enable learners to express genuine evaluations, develop a shared understanding of the characteristics of quality work, select practical steps to improve subsequent performance and to establish another opportunity to observe the learner tackling a similar task, to close the feedback loop;
2. Explicitly setting up learner expectations for quality feedback at the outset, by orientating learners to the purpose of feedback (to help the learner improve on

tasks AND to develop skills in evaluative judgement), setting expectations (such as what a genuine learner self-assessment and interactive dialogue might look like) and propose an outline for the session;

3. Rehearsing practical ways for educators to promote learner opportunities for genuine self-assessment, such as cues to encourage learner participation or pausing to allow sufficient time for learners to respond;
4. Coaching learners and educators on how to discuss aspects of substandard performance and contest opinions through a respectful and honest and dialogue, grounded in curiosity;
5. Further research examining how to create an environment of trust between educators and learners to enable honest appraisal, especially in short-term educational relationships.

5.4 Conclusions

Evaluative judgement is a valuable skill for learners in the classroom and the workplace. Our case study findings support an increasing body of work that suggests that educators provide very few opportunities for learners to exercise judgement within feedback processes. Educators often focus their efforts on delivering their own evaluation of learners' performance and clarifying what good work looks like. Orientating educators and learners to the importance of evaluative judgment and practical ways to enable learners to develop this skill may assist in creating more productive feedback rituals in the classroom, work-based placement, and beyond.

5.5 Chapter 5 Closing discussion

This qualitative study helped to elucidate how educators could work with learners to build their evaluative judgement within verbal feedback encounters. These insights were used to refine items for the Feedback Quality Instrument. A few illustrative examples related to evaluative judgement are provided below; the full details of all refinements to the provisional instrument are described in Chapter 7's supplementary information: Table S7.8.

One of the key findings in this study was that learners' self-assessments were commonly tokenistic, if they occurred at all. Most educators only made cursory attempts at inviting learners to engage in this critical reflection, for learners to compare their own performance with the expected standards and proffer their learning needs. Commonly, educators broadly asked learners for their opinion but learners deflected these invitations to reflect on their work, and educators promptly moved on to describe their own evaluation of the learner's performance. This demonstrated 'feedback as telling' that prioritises 'what the educator does'; the learner is positioned in a passive role, ignoring their volition and agency, and assuming that change in the learner is brought about by the educator.[6]

In contrast, quality feedback from a constructivist perspective, centres on enabling the learner to actively make sense and make use of performance information, arising from their own reflections initially, to instigate improvements.[6,26] In this study, only a few educators were skilled at encouraging learners to share genuine evaluations of their work or employing follow-up probes for the learner to offer reflections deeper than "I think I'm going okay". This seemed to indicate that neither educators nor learners valued the opportunity for the educator to 'observe' the learner's performance analysis and then work collaboratively to enhance it, in other words, to apply similar methods to *learning strategies* as employed for *clinical tasks*.

In particular, these findings led to changes in the two items concerning self-assessment. In the provisional instrument, the item that focused on the educator's invitation was:

Learner self-assessment

The educator asked the learner to identify key similarities and differences between the learner's performance and the target performance.

In the FQI, this was rephased from simply 'asked' to 'encouraged' to emphasise the educator's role to facilitate the learner undertaking a thorough self-assessment:

Learner self-assessment

The educator encouraged the learner to identify key similarities and differences between the learner's performance and the target performance.

The second item was designed to endorse the value of self-assessment. In the provisional instrument, the item was:

Clarify the value of self-assessment

The educator asked what the learner understood about the benefits of self-assessment and helped clarify.

In the FQI, the revised item expanded the educator's responsibility, from simply 'asking' to 'discussing' with the aim of enabling the learner to deepen their appreciation of its value through active participation in dialogue. Further explanatory text was also included, to explicitly specify the benefits of self-evaluation:

Promote the value of self-assessment

The educator promoted the benefits of self-assessment in discussion with the learner.

(Self-assessment provides opportunity for a learner to describe key features of the target performance, analyse their own performance in comparison of the target performance and raise learning needs; this involves practising valuable learning skills such as reflection, evaluative judgement and self-regulated learning).

Of note in this study was that educators invested heavily in clarifying explicit and implicit criteria for quality work in a particular context, often using specific examples from their own practice. Not only did this story-telling bring practice standards to life, but as

learners developed more clarity concerning 'what good looks like', it also dawned on them which aspects of their own performance did not meet these standards, relieving the pressure on the educator to point these out. This reinforced the value of educators working with learners to refine learners' mental schema of the performance criteria, as this made an important contribution to enhancing their evaluative judgement. The findings supported the related item in the provisional instrument, which was listed prior to the item that described the educator explaining their own evaluation of the learner's performance. Hence, the phrasing and position remained the same in the FQI, as no changes were indicated.

Clarify target performance and reasoning

The educator clarified with the learner key features of the target performance and explained the reasoning.

Another key finding in this study, and elaborated further in the qualitative analysis on psychological safety in the next chapter, is the value of dialogue to enable the co-construction of new understandings developed through cycles of inputs and responses. Indeed, fostering evaluative judgement in learners during feedback interactions was not just achieved by skilful prompts or invitations alone. That is, drawing out a learner response, was an important move that was essential to 'set up' the encounter but the greatest benefits arose from the subsequent work educators did in helping to fine tune learners' judgements by confirming or building upon or contesting the learner's viewpoint.

Overall, this study suggested that educators create limited opportunities for learners to make evaluative judgements in feedback. Ironically, at times when learners did raise concerns or ideas about their own performance, educators were swift to move onto other issues that were on their own agenda. This represented a lost opportunity for educators to both engage in dialogue on areas that were important to the learner, and to help calibrate learners' judgement.

As well as helping to refine items, this study, like the study in Chapter 4, served to heighten understanding of contemporary practice in feedback. Understanding the shape of current feedback practice, can help to shape professional development initiatives for educators and learners about how to participate in feedback conversations where both parties contribute to generating knowledge. Specifically, the characteristic language used by educators in their vague invitations to participate, and the common deflection strategies of learners may trigger recognition, and may provide further impetus to build new skills in feedback. Likewise, illustrative examples of productive invitations may be helpful for educators in learning how to enact these instrument items in practice.

In addition, it is noteworthy that this instrument may assist educators to enhance their own evaluative judgement regarding feedback by clarifying 'what constitutes quality feedback practices'. This may encourage educator self-reflection and promote discussion about how these recommendations are best applied by individuals in specific contexts.

The next chapter, Chapter 6, details the second qualitative analysis that explored cultivating psychological safety with learners during feedback discussions.

Chapter 6:

Psychological safety in feedback: What does it look like and how can educators work with learners to foster it?

6.1 Chapter 6 Introduction

In Phase 3, Stage 2 of this research, two qualitative analyses of the rich feedback video data were performed. The first analysis was described in Chapter 5 and investigated how educators can support learners in building their evaluative judgement during feedback conversations. The second qualitative analysis of the feedback video data is described in this chapter and focused on exploring how educators can promote psychological safety in collaboration with learners.

The importance of the educator-learner relationship, trust, a learner's belief about whether an educator has their 'best interests at heart' and psychological safety were discussed in the paper reported in Chapter 3. This reflects the powerful interpersonal dimension of face-to-face feedback discussions. In particular, these notions underpinned *Item 11: The educator showed respect and support for the learner*. However, it remained somewhat unclear 'what this looked like' in practice. In other words, there was still a gap in how such a desirable target to 'show respect and support' could be operationalised by educators.

The article that follows on psychological safety was built on two platforms: analysis of video data and extensive reading, particularly articles on psychological safety and allied concepts beyond feedback by renowned experts including Amy Edmondson, psychologist who investigated psychological safety, learning and performance in healthcare teams[192,193]; Lisa Rosenbaum, cardiologist and National Correspondent for the New England Journal of Medicine who instigated a series on psychological safety[194]; William Bynum, family physician and researcher in psychological safety, personal connection, shame and guilt[31,195]; Jenny Rudolph, psychologist, and Walter Eppich, paediatric emergency physician, and researchers in fostering 'a safe container' in

healthcare simulation-based education[196,197]; and Jonathan Silverman, family physician and researcher on doctor-patient communication skills.[198]

Using the video data in this study, we could only infer a learner's perception or experience of psychological safety. The insights gained from this extensive reading, especially the work of Edmondson, underpinned the idea to look for markers for when learners experienced psychological safety. That is, we used times when learners demonstrated vulnerability and progressively undertook productive learning behaviours such as asking questions, discussing mistakes or giving an opinion, as signals that the learner felt some level of psychological safety. Then corresponding educator behaviours surrounding these learner behaviours were analysed. Using these data, themes were developed that described ways that educators appeared to foster psychological safety with learners. This work is described in detail in the following article.

This publication won the 2020 Melbourne Medical School Annual Publication Prize (students), University of Melbourne, and was the focus of a podcast interview for the journal Medical Education (Kevin Eva and Christina Johnson, June 2020). Link here

Data analysis

The details regarding my contribution to the data analysis presented in the following chapter are detailed here. The video data were interrogated using thematic analysis. I independently read and open-coded a sub-set of transcripts in parallel with the core research team (EKM and JLK). We met on multiple occasions to discuss our independent analysis and arrived at a coding framework (see Supplementary information: Table S6.1). Once our coding framework was collectively established, I led the thematic analysis of subsequent transcripts and utilised NVivo software (NVivo Windows 12.5.0, QSR International). Codes were defined to aid understanding and further contribute to the transparency of the analysis process. New themes were added to the coding framework in consultation with the research team, if findings were not reflected by the codes within the initial coding framework. Memo notes were kept throughout the analysis process by

the primary researcher (CEJ) to document new questions arising, or links to existing literature.

Psychological safety in feedback: What does it look like and how can educators work with learners to foster it?

This chapter is the study published in Medical Education 2020;54:559–570. (Lightly edited for headings and formatting consistency). Published June 2020.

Christina E. Johnson[1,2]; Jennifer L. Keating[3]; Elizabeth K. Molloy[2]

[1] Monash Doctors Education, Monash Health, Clayton, Victoria, Australia

[2] Department of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia

[3] Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine Nursing and Health Science, Monash University, Melbourne, Victoria, Australia

Keywords: Psychological safety, Feedback, Formative feedback, Effective feedback, Health professions education, Workplace learning, Professional development

Funding: none

Declaration of interest: none

6.2 Abstract

Context: Feedback conversations play a central role in health professions workplace learning. However, learners face a dilemma: if they engage in productive learning behaviours (such as asking questions, raising difficulties, offering opinions or contesting ideas), they risk exposing their limitations or offending the educator. This highlights the importance of psychological safety in encouraging learners to candidly engage in interactive dialogue and the co-construction of knowledge. Previous research has recommended that building safety, trust or an educational alliance is key to productive feedback encounters. Yet it is unclear how to translate this into practice. Hence our research question was: what does psychological safety look like in workplace feedback and how can educators work with learners to foster it?

Methods: We analysed 36 videos of routine formal feedback episodes in clinical practice involving diverse health professionals. A psychologically safe learning environment was inferred when learners progressively disclosed information and engaged in productive learning behaviours during the conversation. We used thematic analysis to identify associated educator strategies which seemed to promote psychological safety.

Results: Four themes were identified: *'Setting the scene for dialogue and candour'*, *'Educator as ally'*, *'A continuing improvement orientation'* and *'Encouraging interactive dialogue'*. Educators approaches captured within these themes, seemed to foster a psychologically safe environment by conveying a focus on learning, and demonstrating respect and support to learners.

Conclusions: This study builds on claims regarding the importance of psychological safety in feedback by clarifying what psychological safety in workplace feedback conversations might look like and identifying associated educator approaches. The results may offer educators practical ways they could work with learners to encourage candid dialogue focused on improving performance.

6.3 Background

Feedback plays a central role in workplace learning in health professions education.[199-202] Commonly in clinical practice, this involves a senior clinician ('educator') and a junior clinician ('learner') discussing the learner's performance. This may be in formal settings, such as workplace-based assessments or end-of-attachment appraisals, or embedded informal discussions, such as after a clinic or during wound-closure in theatre. The aim of feedback is to assist learners to understand what 'quality work' looks like and how their work compares with these standards, and to use this performance information to implement practical strategies to improve performance.[1,18,19,113,172,203]

To achieve these aims, there is consensus in contemporary feedback literature, underpinned by a social constructivist stance, that learners' active engagement is essential. Learning conversations offer opportunities for learners to refine their understanding and contribute to shaping co-constructed ideas.[19,28] For example, learners might articulate a thoughtful self-assessment by comparing their own performance with the desired performance. This process of reflection and evaluative judgement engenders valuable learning in itself (as well as offering further avenues for learning).[20,172,204] In addition, dialogue between an educator and learner offers the opportunity for a catalytic reaction, as their different perspectives interact, to co-create new insights and strategies uniquely tailored to assist the learner.[166,167,205,206] All this requires a learner to candidly reveal their learning needs and negotiate the design of an improvement strategy that will work for them.[207] On the other hand, if a learner does not reveal the difficulties they have experienced, ask questions, propose their own ideas (even if partial or flawed) or contest an idea, then an educator's advice may not be sufficiently relevant, useful or acceptable for the learner to put into practice.[166,208]

However, despite the espoused importance of learner participation in optimising feedback outcomes, this does not appear to have been translated into routine practice. Observational studies have reported that feedback commonly comprises 'one-way' information provision, dominated by the educator.[57,114,115] Learners often report that

educators' comments are not specific, clear or actionable and they struggle to 'make sense' and 'make use' of it.[11,26,100,119,209]

To counter this lack of learner involvement, there has been increasing focus in the literature on the value of the educator-learner relationship in creating safe spaces to allow for learner contribution. In health professions education, Ende called for the learner and educator to work 'as allies'.[3] The R2C2 (relationship, reaction, content, coaching) feedback model includes 'relationship building' as the first phase, with the goal to 'build a positive relationship based on trust and respect'.[167,168,210] The concept of an 'educational alliance' has also been proposed, derived from the 'therapeutic alliance' in psychotherapy. A strong therapeutic alliance is said to exist when a client believes the therapist shares with them a positive bond, the same therapeutic goals and strategies for working towards them. Junior doctors, who felt a strong educational alliance with their supervisors, reported seeking feedback, being open and honest during discussions and implementing planned strategies.[55,99] In higher education, Carless explored the role of trust in dialogic feedback.[29] He used a definition by Tschannen-Moran that trust is one's willingness to be vulnerable to another based on an investment of faith that the other is open, reliable, honest, benevolent, and competent.[211] Carless asserted that trust was essential if learners were to take risks and fully engage in challenging learning activities that may reveal their weaknesses.

We propose that psychological safety is an overarching construct that ties these concepts together (working as allies, the educator-learner relationship and trust), which has a powerful influence on engagement and learning. Drawing on Edmondson's work within psychology, psychological safety may be defined as "a shared belief that the [educator-learner] relationship is safe for interpersonal risk taking", which creates "a sense of confidence that [the educator] will not embarrass, reject or punish [the learner]…due to mutual respect and trust".[193(p354)] Edmondson's research exploring interpersonal dynamics and healthcare outcomes found that in nursing teams with high psychological safety, productive learning behaviours were commonly seen, such as asking questions, discussing mistakes, seeking help or trialling new ideas. The

environment was characterised by respect for others' expertise; trust that people had good intentions and were invested in others' success; and interest, acceptance and care for people as individuals.[193] Team leaders held high standards and were continually seeking to improve. Yet they were also supportive, coached their team, minimised hierarchy, were not defensive when challenged, and role modelled learning behaviours.[193]

The effect of psychological safety on learning and performance has been demonstrated in other areas of healthcare.[192,194,212,213] In simulation-based education, the importance of psychological safety is particularly emphasised; it is explicitly labelled as a goal at the outset and both educators and learners commit to ground rules that 'realise it' in practice.[196,197,214] In the area of doctor-patient communication, psychological safety promotes greater patient involvement and free flow of information between doctor and patient, which results in more accurate assessments, shared decision making, resolution of uncertainties, implementation of plans and improved clinical outcomes.[215]

Applying this to feedback interactions, it may be that psychological safety is a key mediator encouraging or deterring learners' participation. When learners use productive learning behaviours, they may reveal previously undetected limitations and trigger adverse responses from the educator, for example a disparaging remark, lower grading or exclusion from a group of the 'favoured few'.[31,193,210] Potential threats are amplified when educators are senior health professionals with power over learners' progress, such as having responsibilities for summative assessments or professional references.[216,217] Medicine, in particular, has a powerful hierarchy.[218,219] Hence a learner may decide to keep quiet, or, at least, to control the risk by regulating their exposure. The key question for any learner contemplating what to volunteer during a feedback conversation is: 'What is the likelihood that I will be respected, accepted and assisted or the opposite, that I will be humiliated, reprimanded or judged as inept?'[196,210,218,219]

If educators could work with learners to maximise psychological safety, feedback conversations might be transformed. Learners might honestly reflect on their performance, explain their reasoning, reveal their difficulties, ask questions, propose

their own ideas for improvement or contest educators' comments. If these learner contributions interacted with valuable educator inputs, the resulting co-construction of knowledge could substantially enhance learning outcomes. While this construct of psychological safety has been the focus of research in other contexts, the manifestation of psychological safety in workplace feedback discussions has been under-examined. Hence, our research question was: What does psychological safety look like in workplace feedback and how can educators work with learners to foster it? Our definition of feedback, and the theoretical positioning of this research, sits within social constructivism with acknowledgement of the interdependence of social processes in knowledge generation.[220]

6.4 Method

Collection of videos of feedback episodes

As part of a broader program of research investigating feedback practices, we collected self-recorded videos of formal face-to-face feedback episodes in routine clinical practice. Participants involved health professional educator-learner pairs who were working together, within a major Australian hospital network.[57] During recruitment, diversity was sought by rolling advertising to potential participants, with consideration of key factors including health profession and specialty, gender and supervisor experience (educators) or training level (learners). All participants gave written informed consent. Ethics approval for this study was obtained from the health service and university human research ethics committees (References: 15233L and CF13/1912-2013001005 respectively).

Videos were professionally transcribed verbatim and one researcher checked accuracy by watching the videos while reading transcripts (CEJ). While viewing videos, initial impressions and observations were recorded as memos, and educators' practice was analysed using the Feedback Quality Instrument (provisional version). This instrument is designed to assist educators in facilitating high quality, learner-centred feedback, aligning with a social constructivist paradigm. The items describe observable educator behaviours considered to engage, motivate and enable a learner to improve. Items were

developed through a process involving an extensive literature review to identify information that substantiated ways educators could promote beneficial learner outcomes and a Delphi process with a panel of experts.[52,57] During further testing and refining of the feedback instrument, the gap in the literature on psychological safety became apparent (as described in the Introduction), which was the inspiration for this study. We chose to focus on the educator's role because of their power to steer feedback interactions.

Data collection and analysis for this study

In this study we looked for signs of psychological safety and associated educator approaches which seemed to help achieve the conditions that enabled candid learner contributions. We reasoned that learners' propensity to be forthcoming and participate in 'risks for the sake of learning' were indicators of learner psychological safety. Hence, we searched for occasions during the feedback sessions when learners repeatedly undertook productive learning behaviours, particularly when learners became progressively more candid about their learning needs. Examples of these learning behaviours included sharing information, describing a difficulty, identifying their own sub-standard performance, naming a learning need, asking for help, asking a question or proposing a solution.[193] We then looked for corresponding educator behaviours that seemed to promote this form of engagement. We also looked for educator behaviours previously proposed to promote psychological safety in clinical and simulation-based conversations.[193,196-198]

Transcripts were inductively coded according to principles of thematic analysis, guided by our research question focussing on psychological safety and sensitised by our theoretical framework of social constructivism.[127] A total of ten transcripts of videos rated most highly using the feedback instrument (reflecting a greater number of quality educator behaviours observed) were used to develop a coding framework, as these seemed most likely to involve a psychologically safe environment. First, four transcripts were open coded independently by each researcher (CEJ, JLK, EKM), then researchers met to discuss and create an initial set of codes. Following this, a further six transcripts

were coded independently then discussed in pairs (three by CEJ and EKM; three by CEJ and JLK), using and modifying the initial coding framework. Further team discussions led to consensus on a final coding framework with operational definitions and codes with similar properties grouped together. It included markers suggesting when learners felt psychologically safe and when they did not, noting corresponding educator behaviours.

The lead researcher (CEJ) then re-watched each video and re-coded all transcripts using the analytical framework utilising NVivo 12.5 (NVivo QRS International Pyt Ltd, Vic, Australia). Interpretation and assimilation of codes, with descriptive quotes, into themes occurred at three separate research team meetings, with the final key themes agreed by consensus. In each meeting, a constant comparison process was utilised to identify consistencies and differences in coding between researchers.[221] Repeated examination of the transcripts and coded data increased the likelihood that themes were grounded in the data.

This study was approved by the Human Research Ethics Committee at Monash Health (Reference 15233L) and Monash University (Reference CF13/1912-2013001005). Written informed consent was obtained by all participants.

Theoretical framing

Social constructivism is the theoretical framework underpinning our interpretation of productive feedback interactions. According to social constructivism, knowledge is created during the process of people interacting, particularly through discussion; hence the learning is specific and intrinsically linked to the context and people involved.[43] As effective feedback depends on a learner implementing desired changes, it is crucial for the learner to play an active role in understanding and contributing to the discussion. Hence, although we focused on educators' approaches to helping establish psychological safety, we used learners' behaviours to signpost periods during the feedback interaction that we wanted to examine more closely. In our analysis of psychological safety using observational methods, we were aware of the risks of interpreting a phenomenon that is 'held' or 'experienced' by individuals. However, we

viewed evidence of learner risk-taking and co-construction of knowledge as markers of an interaction that was enabled (or limited) by psychological safety.

Reflexivity

Reflexivity, the influence of the researchers' own perspectives, beliefs and experiences, on the interpretation of the data, was explicitly recognised and discussed. CEJ is a physician, medical educator, and researcher with a special interest in feedback, who works at the health service involved in this study. JLK and EKM are both physiotherapists with extensive expertise as health professions educators and researchers in feedback and workplace based assessment. Their employment is university-based and not linked with the health service in this study.

6.5 Results

A total of 36 videos of feedback sessions were collected. Participants included 34 educators (26 medical across 9 specialties, 4 nursing, and 4 physiotherapy professionals with a range of supervisor experience; 16 (47.1%) were male) and 35 learners (students and clinicians with a range of experience; 12 (34.3%) were male). Feedback was related to a specific task in 25 (69.4%) videos and a mid/end-of-attachment appraisal in 11 (30.6%) videos. Most of these feedback occasions were formative in nature but some served as summative contributions to longitudinal evaluations.

Educators used a variety of approaches to promote psychological safety. This included educators' setting the scene for a candid feedback interaction, positioning themselves as allies alongside the learner, focusing on continuing improvement and encouraging interactive dialogue, which signalled to learners that honest contributions were expected and valued. Four key themes are presented below with illustrative quotes.

Theme 1: Setting the scene for dialogue and candour

When feedback sessions began, many educators simply announced the topic for the session, for example: '*So we're going to talk about the surgery today'* [V4] and then moved swiftly onto performance evaluation.

A few educators did more to set up the interaction. Here, an educator highlighted the aim of the session and the importance of the learner's input:

> *Educator: It's midway through the rotation, so it's a good opportunity to…touch base and see where things are at, see how you feel things are going, and…make sure things are on track. So, from your perspective, how do you feel the…rotation is going so far?*
>
> *Learner: For me, the rotation has been very amazing. Psychiatry has always been my weakest area during my medical training, and that's one of the main reasons that I have requested to come here. [I] don't normally do a comprehensive psychiatric assessment…so I find that it is quite challenging. But, on the other side, I'm learning combining mental health along with the medical issues…I find that…very rewarding. [V19]*

In this next example, the educator proposed their plan for the session, so the learner knew what to expect. Yet the most outstanding feature was the educator's explicit invitation for the learner to contest their comments or ask questions, thereby encouraging productive learning behaviours:

> *We've just gone through a clinical scenario and what I'd like to do is talk about a couple of aspects of your performance where you've done some really good skills that are worth continuing with and identify…a couple of areas that you can work on, over the next little while, to improve your performance...feel free to stop me at any point in time if I say things that you don't agree with or that you need clarification on. [V3]*

However, educators did not comprehensively set the scene for a quality feedback interaction by highlighting the purpose (to help learners improve), proposing a plan for the session (to analyse their performance and develop a learning plan), setting expectations for an interactive dialogue focussed on learning and clarifying the time available. Even when educators did include some of these components, none sought learners' input to negotiate any aspects of the session structure.

Theme 2: Educator as ally

Educators commonly positioned themselves as an ally, working alongside the learner. They offered support and assistance, to facilitate learner development and actively tried to reduce the power gap.

Offering support and assistance

Educators frequently used the collective pronoun 'we', suggesting solidarity with the learner, and expressed empathy for the struggles that learners faced.

> *Educator: So, arterial lines: like I said before, they're the things that **we** stuff up, that **we** fail at the most often….They're difficult, they're fiddly…and…often **we**…take multiple attempts to get [them in]…Saying that, what can you do to be better at them? Have you used the [tradename] kits?*
>
> *Learner: Yes, I have used them before…I thought…it would actually…be a bit too big for [the patient's blood] vessel because she was quite dainty, and…they seem a bit chunkier…and with smaller vessels, I've struggled to get them in. So that's why I didn't initially go for that.*
>
> *Educator: No, fair enough…What's your default?...The advantage of [these kits] is, as soon as you get a flashback [of blood once you're in the artery]…you can just put the [guide] wire in…So…have a think about it…I think they're handy but obviously not everyone uses them. [V28]*

It was very common for educators to offer advice to help learners improve. Often these were useful practice tips (as above). In addition, some educators addressed meta-cognitive processes, for example by suggesting strategies to make the most of learning opportunities:

> *Don't be afraid to put your hand up and say, "Can I come and watch you do that?"...if it's…a learning opportunity that you need…I know that's hard…when it's busy…but it might be that you offer to do something on their behalf. [V36]*

Reducing the power gap

The literature describes how the hierarchy within the health professions may exacerbate a learner's fear of negative consequences if they expose their limitations or contest ideas. We found some educators demonstrated humility and tried to reduce the power gap. A few educators did this by revealing their own vulnerabilities, such as admitting they had made mistakes, did not know everything or were still learning:

> *So there's always things to learn, and I'm still learning [V17]*

Educators actively shared power by showing respect for learners' opinions or not assuming their own judgement or advice was indisputable. Here an educator indicated awareness that their evaluations may not accurately reflect the learner's capability due to limited observations:

> *Like I've said before, it's hard to differentiate between 'always demonstrates' and 'mostly demonstrates' [rating options on the assessment form] when I'm not working with you every day. [V36]*

Similarly, an educator specifically explored how their presence may have affected the learner's performance:

> *Educator: So we've just gone through the mini-CEX [mini-clinical examination examination] process. You mentioned you haven't done one of those before. How did you find it?*
>
> *Learner: I thought it was good. It's different having somebody there…As a student…you often have somebody in the room but…when you work for a while, and don't have someone in the room…it's a little bit, ah, daunting at first but you get used to it.*
>
> *Educator: Do you think you did things differently because I was standing there watching?*
>
> *Learner: …because I was being assessed, it kept running through my mind whether I'd missed something, which I wouldn't normally do. I'd do that as I was writing notes and then go back and check with the patient. [V9]*

Some educators used phrases such as 'I suggest' or 'I would recommend' when they offered ideas, rather than 'you need to' or 'you must', and explained their reasoning. Educators also checked learners' thoughts about their judgements or recommendations:

> *Educator: When we're doing consultant ward rounds, I like the way you present what's happened to the patient [and] the issues but I'd like more of a management plan as well… Making sure that you're making those decisions and…taking ownership of the patient…Just a few things that you can work on and become even better.*
>
> *Does that sound fair?*
>
> *Learner: Yeah, that's good….I suppose, when you're a reg [registrar], I feel like, "Oh, [the consultant]'s here to fix all the patients"…It's always…how I have been…just to present the issues.*
>
> *Educator: But I think you…can step up…If there's anything that you're not quite sure about, you can say, "Hm, I'm not quite sure, but I think this is the way that I would go."… So even if you're not 100% sure, you suggest something. I mean, when you become a consultant, …you may not know all the right answers but it's all about making a [reasonable] decision.*
>
> *Learner: Yeah, all right, sounds good. I'll work on that. [V1]*

In summary, many educators used strategies to convey their desire to be allies with learners. Educators' helpful, caring, humble and respectful responses (and the absence of punitive ones) seemed to build trust. This was evidenced by learners frequently responding to these safety signals by revealing more of their perspective, as demonstrated in the passages above. Learners were monitoring educator cues within the feedback session, which appeared to influence what they chose to contribute subsequently.

Theme 3: A continuing improvement orientation

We observed ways educators attempted to create a psychologically safe environment when dealing with learner's inexperience, mistakes and skill gaps. This approach was characterised by a focus on continuous improvement.

It's OK, gaps and mistakes are expected when learning

Some educators expressed that skill gaps or errors were expected during skill development:

> *It's okay to make mistakes, and stumble,…because…you're a student and the point is to learn. [V10]*

A common strategy was for educators to normalise a learner's mistake, by saying it was common or admitting they had made the same mistake when learning the skill. Here a surgeon gives some technical advice:

> *You need to go…directly down, rather than into it, which is a common mistake. I did it when I first started and everyone does it. [V24]*

Educators often mentioned the learner's experience or training level during performance evaluations. This suggested that educators aligned their expectations accordingly, commending fledgling skills and accepting deficiencies early in the learning journey:

> *When you came in, your neuro assessment skills…weren't particularly strong. But…we need to take into account the fact that you haven't actually covered that at [university] yet. [V10]*

At times educators worked with learners to reduce excessively high learner expectations:

> *Learner: …sometimes it gets too busy…one discharge, one admission… [I would like] to learn how to manage all the things on time…The nurses…sometimes get really, really busy and they manage really, really well…*

Educator: …You guys come in for two weeks and then you're gone again… don't stress too much…because you just don't spend enough time in the ward to have really good time management. [V10]

Educator focuses on learning strategies

Some educators commended learners' effort, strategies and progress, thereby privileging ongoing development over validating performance:

Educator: I think your critical thinking is really good…From where you started to where you are at now, that ability to stop and think, "Okay, I'm doing this; this is why I'm doing it", is really good…What sort of strategies have you put in place to get there…?

Learner: I think [after an experience]…which I'm not familiar with, going to someone or going back home and…reflecting and then going to…our study day notes, just to scrub up on things.

Educator: You can..definitely see that in your work; that you've gone home, looked things up …You're a very keen learner, which has obviously helped in your practice. [V36]

In summary, some educators focussed on how learners could progressively develop their skills. They took account of learners' experience, provided commentary which implied that gaps and errors were an integral part of the learning process and concentrated on optimising learning strategies that could assist with ongoing development.

Theme 4: Encouraging interactive dialogue

We found that interactive dialogue, when both perspectives were voiced and responses were related to the other's comments, formed the basis for co-construction of tailored improvement strategies. This went beyond turn-taking; there was evidence of both educators and learners modifying their views, as new information was considered.

Sharing information

First, we looked at how information was shared, offering the opportunity for each party to hear and consider the other's perspective. We found that educators routinely expressed their thoughts and opinions. In contrast, learners typically only spoke when educators asked them questions and even then, their initial answers tended to be brief and superficial.

> *Educator: Why don't you start off by telling me how you went?*
>
> *Learner: I thought it was okay…nothing went terribly wrong. [V11]*

A few educators made additional efforts to gain a greater understanding of the learner's perspective by exploring it more thoroughly. The following excerpt concerned a learner's assessment of an elderly man, accompanied by his daughter, in the emergency department:

> *Educator: If you had your time again, is there anything that you would have done differently?*
>
> *Learner: Uhm.. not that I can think of…[long pause, uninterrupted by the educator]*
>
> > *I think uugh...maybe, just clarify a few more things with the patient themselves...I spoke quite a lot to the daughter.*
>
> *Educator: Mhm, mhm.*
>
> *Learner: I got a little bit of information out of the patient, but he…seemed to be referring to his daughter, anyway, for the history, which is why I turned to…her for a bit more information.*
>
> *Educator: Yes.*
>
> *Learner: Uhm, so maybe… starting off with the patient a bit more…and then just checking in with the… daughter for further information.*
>
> *Educator: Mm.*
>
> *Learner: Yeah. [learner indicates she has finished talking]*

Educator: It's always difficult when there's family in the room that know more; they keep wanting to interject but you're trying to build a relationship with the patient [first] [V9]

Here, the educator used open-ended questions and demonstrated attentive listening. These skills include remaining silent when the learner paused, to allow them to think and then continue talking; encouraging (sounds that convey, 'I'm listening, go on' such as 'hmm' or 'yes'); summarising and echoing. After an initial deflection, the learner reflected on events, revealed their difficulties and proposed a potential solution (all valuable learning behaviours). The educator's comments were characterised by empathy and assistance, not criticism, thereby promoting psychological safety.

In particular, learners typically only revealed difficulties when they were specifically asked about them:

Educator: What are some of the things that you're finding difficult…on the ward?

Learner: I still feel like, with the patient load,…that I haven't really…gotten on top of that yet…There's still things…that I'd like to do that I just haven't got the time to do. [V1]

Alternatively, some educators cited a specific instance and then asked learners for their comments. This example comes later in the emergency department interaction:

Educator: When you entered the room, tell me about the way that you introduced yourself…to the daughter.

Learner: [I] made the mistake of - I knew she was in the room, but I didn't - I should have initially…asked who she was. [V9]

Responding to the other's comments

Next, we explored how participants responded to the other's comments, to complete communication loops. Here, a student was discussing their assessment of a patient presenting with paranoia:

Educator: What else do you think you could have improved on?

Learner: Probably handling her getting a little bit upset or…completely shutting down…I had a feeling she might, regarding that topic…Do you automatically just stop everything and go, "Okay we'll come back to it later", or can you get past that?

Educator: That's the key information that you really want to find out…Now obviously you don't want to...make it escalate and get aggressive…In situations...where it is a touchy subject, you may sit back and, just for three or four seconds, wait and see whether they carry on with the conversation.

…Did you realise how it took a little bit of time for us to actually see her real presentation? So, like in the first…10 minutes, she kept it together pretty well…then it…started - she was getting irritable, she wanted to end the conversation -

Learner: Was it 15 minutes, really? It didn't feel that long.

Educator: Yeah…that's why you need to…keep the conversation going for a longer period of time…getting them to…tap in a little bit more…

Learner: That makes sense because I found the risk assessments that we do..each shift…can be…very abrupt… In those five minutes, not everyone is open enough to -

Educator: Correct.

Learner: … I hear later on [that] they've got paranoid thoughts or…delusions but they never actually showed that when I was…talking to them. [V35]

Notably, the learner directly asked the educator for help; this bold move was unique within our data set. Generally when learners mentioned a difficulty, they appeared to assume this was a sufficient hint and did not explicitly ask for assistance (some educators responded but others did not).

As the speaker's role alternated, it allowed each participant to frequently contribute to the shared mental construction zone, in which ideas (or gaps) were clarified, tested and developed. The educator endorsed the learner's dilemma, offered a solution and then focused back on the learner. Throughout, the learner had a central role in the discourse,

with the educator taking their cues from the learner and tailoring comments accordingly.

However, such interactive dialogue was not common. Commonly educators ignored learners' comments and pursued their own agenda. They typically only asked, 'Do you have any questions?' at the conclusion of the session. Learners habitually acknowledged educators' comments by saying 'yes' but little input beyond this, and rarely sought a response from educators to their comments.

In particular, views were scarcely contested. In our data, no learner overtly disagreed with an educator's remarks. When educators did express a different opinion, mostly they were reassuring learners that their work was satisfactory, contrary to the learner's concerns. In this unique example, an educator counters a learner's view:

> *Educator: …I would encourage you to…[propose] how…to settle this...You had a [young] boy and…parents, who were very strong in their opinions and their wishes on things…In terms of your management plan: I think you knew what you wanted to do but you felt unsure because…of [how the parents might react].*
>
> *Learner: The moment I was examining the child...and the mother…strongly voiced her concerns…I immediately…went, "Maybe I should have a consultant here before I even suggest [the required procedure] to the parents"…In paediatrics, you tend to take a step back when a parent raises their voice because you're worried, maybe they are right, because occasionally they are right when they raise their voice.*
>
> *Educator: …It's a difficult balance, isn't it, between having conviction in your own approach and being open to the concerns of the parents, which we always need to listen to, even if they're not necessarily right or helpful [for the child]? That's not just an experience of being a little junior in the team, that's an experience you get all the way through your career…You will have parents who have their own ideas about… how this needs to be managed, and you're beginning to develop skills on how to engage them in that process and the fear that usually underlies that behaviour, as well as making sure that the priorities for the patient's management are still followed…*

Learner: Hmm, definitely. The other thing…is - I'm worried that sometimes when, as a junior you're overconfident, given you don't know much, you go into a room thinking, "Yep, I can handle this situation" but over-confidence…can lead to more harm. Though I think I've always been in the under-confident side, like, "Always seek help before you put something in or do something," because after that, back tracking won't always help you.

Educator: No, you're absolutely right and we always need to be cognisant of safety…You know most of the time quite clearly what needs to happen clinically… having a bit of confidence…and saying [to the parents], "Look, this is what I think I need to do but I will go and clarify that with my consultant because [your child]'s obviously in distress and you've got some strong ideas about what needs to happen." [V3]

The educator clearly repositioned the parent's views as paramount, while empathising and normalising the predicament this created for the learner. The educator continued by sharing their clinical insights into this challenge. This response seemed to promote psychological safety as the learner then candidly revealed what had held them back from initiating the management plan. This gave the educator the chance to more accurately target their advice to address the specific problem the learner was grappling with.

In summary, we identified examples of interactive dialogue in which learners openly shared their mindset and participants responded to each other's comments. This building on the thoughts of others was distinct from 'turn-taking' as a proxy for 'two-way interaction'. However, this scenario was uncommon. Learners tended to speak only when invited. Therefore, learners' difficulties, reasoning, questions, opinions or ideas were mostly never voiced, and hence could not be responded to. However, a few educators used exploratory questions and attentive listening; this seemed to sustain the invitation for the learner to speak and convey that the educator was interested in and valued the learner's input. This approach appeared to contribute to building psychological safety and encouraged candour from learners.

However, mostly communication loops were 'left hanging' incomplete, when participants did not respond to the other's contributions. Educators did not routinely

respond to learners' comments. These were lost opportunities to promote psychological safety by reducing uncertainty in the learner's mind regarding the educator's opinion on what the learner had said, or to offer support and assistance. In addition, both parties appeared reluctant to debate or disagree. Hence, most chances to find out more, clarify, challenge or modify discussion points during the conversation were not utilised. Nevertheless, on the few occasions when both participants engaged in an interactive discussion, we saw an evolution in thinking on both sides resulting in explicitly shared understandings (common ground) and co-development of tailored improvement strategies. By encouraging the learner to articulate exactly what they were struggling to do, the educator was able to customise their advice.

Feedback interaction illustrating all four themes

For an extended excerpt showcasing what psychological safety could look like and how an educator might work with a learner to foster it, see supplementary information: S6.2.

6.6 Discussion

We have presented our analysis under four themes that describe how educators worked with learners to create psychological safety in verbal feedback sessions: *Setting the scene for dialogue and candour; Educator as ally; A continuing improvement orientation;* and *Encouraging interactive dialogue*. Despite the literature advocating that aspects of psychological safety such as 'trust', relationship building' and 'educational alliance' are crucial for effective feedback, there is little empirical evidence of what this looks like in workplace feedback conversations.[26,29,99,168] This is the gap we have attempted to address, using empirical research, to describe ways that educators worked with learners to foster psychological safety in formal verbal feedback sessions. However, the approaches presented here are neither prescriptive nor exhaustive.

Importantly, there were indications of an undercurrent or a 'meta feedback loop' for the learner, in which educators' reactions promoted or inhibited learners' openness and engagement in learning within the conversation. Indeed, this highlights the value of an educator setting the scene for dialogue and candour at the start of a feedback session (although this explicit invitation requires ongoing support throughout the session by

what the educator *does*). Learners appeared to be making moment-by-moment assessments about psychological safety and the risk-to-benefit analysis of what to say. When educators offered support and assistance (not criticism), we saw learners progressively engage more candidly in learning activities like explaining details, offering their opinion, broaching drawbacks in educators' recommendations or suggesting alternatives. In other words, learners were dynamically reading cues from the educator, to determine whether they should advance or retreat. When a learner 'dipped their toe in', where the response fostered psychological safety, the learner boldly took a step deeper. However, it is important to note that in many of the feedback sessions in this study, educators dominated the conversation and when a learner made an initial comment, this was ignored. This closely mirrors research into doctor-patient communication, in which doctors' responses strongly influenced how much patients disclosed during consultations.[215,222,223] Frequently doctors ignored patients' initial hints, which suppressed further elaboration and cooperation.[224]

This pattern reinforces the importance of assisting educators to understand the value of learner-centred feedback interactions, and offering practical ideas on how they could operationalise this construct in clinical practice in collaboration with learners. Clinical educators hold a high degree of structural power over learners (both students and junior clinical staff) as control over training progression is commonly embedded in educators' roles.[207,225] Helping educators to be more conscious of the impact of their invitations and responses (or lack thereof) on learners' propensity to take productive intellectual risks, may motivate educators to change their practice. Having linguistic examples of how to elicit and respond to another's agenda may assist educators to enact changes and have a positive impact on feedback rituals in clinical practice.

Our findings overall reinforce previous research across the fields of higher education, health professions education and communication, which have reported that learning behaviours are fostered by trust, respect, empathy, seeking to understand the learners' perspective and an attitude that accepts mistakes as inevitable during learning.[29,198,210] In simulation-based education, a set of 'promising practices' have been proposed for

creating a safe context for learning, derived from a literature review and the team's expertise.[196,197] These included clarifying expectations; attending to logistical details; showing respect and curiosity regarding participants' perspectives; understanding that mistakes occur in challenging learning environments, focusing on assisting learners to enhance their skills and maintaining positive regard for them. Some of the themes we identified are consistent with these experts' insights into the phenomenon in simulation-based contexts, but our examples of practical behaviours and dialogue offer options that educators wanting to build psychological safety, could tailor for specific contexts during their own feedback encounters.

The theme *'Setting the scene for dialogue and candour'* examined how educators set the tone for the session upfront. In our study, this was not done well. When educators indicate their intention to assist and highlight the value of learners' contributions from the outset, this endorses interactive dialogue and enables learners to better position themselves to actively participate. It is likely that some of the educator-learner pairs did not know each other well. In this context, being explicit about expectations is likely to be even more important as routines have not been established. By creating transparency and offering some degree of control about what is going to happen and what is expected, learners are likely to feel less anxious. This is important for learning outcomes as excessive anxiety distracts attention, interferes with complex thinking and impairs memory.[159,226] This theme aligns closely with guidelines for simulation-based education and doctor-patient communication skills, which outline the importance of setting the tone for a collaborative partnership from the start.[196,197,227]

The next theme *'Educator as ally'* identified how educators offered support and assistance to learners and purposefully attempted to reduce the power imbalance. These strategies seem to foster psychological safety by increasing the likelihood that 'taking a chance on learning' would reap benefits for learners. Bearman and Molloy used the phrase 'intellectual streaking' to describe instances when educators 'exposed' their own limitations to learners, with the potential to build trust and reduce the power imbalance through this gesture of reciprocal vulnerability.[228] In a subsequent article, the

same authors explored the 'vulnerability-credibility' tension where educators revealing their uncertainties regarding complex clinical problems to facilitate learning, were also trying to maintain their standing as experts. They advocated for the value of "building a culture that acknowledges fallibility" - as errors are central to learning throughout life at work - "rather than honouring perfectionism".[207,229]

The theme *'A continuing improvement orientation'* aligns with a 'growth mindset', as described by Dweck.[157] This mindset privileges ongoing learning and focuses on developing skills by seeking out challenging tasks and applying effective strategies, effort and persistence. Mistakes and knowledge gaps are seen as opportunities to learn, not cause for covering up by a learner or criticism from an educator. This offers an approach to combine high professional standards with an understanding that mistakes or gaps are expected while learning.[53]

The final theme *'Encouraging interactive dialogue'* describes ways that educators facilitated discussions in which participants built on each other's contributions. This led to the creation of common ground (explicitly shared understandings), and the co-construction of development strategies targeted to address the learner's difficulty. This key finding aligns with recommendations to 'build a history' by equally prioritising the patient's and doctor's perspectives during a consultation.[198,222] This has been likened to "two writers collaborating on a manuscript…until both are satisfied".[223] This paradigm unique perspectives has been shown to result in better patient trust, satisfaction and clinical outcomes, probably related to improved adherence to plans developed collaboratively.[223] However, in our study, and other studies of feedback,[230] this was not commonly seen.

Strengths and weaknesses

Our study analysed 36 videos of authentic formal verbal feedback sessions during routine clinical practice with diversity of health professional participants across specialty, experience and gender. We examined the context in which learners increasingly disclosed information and engaged in learning activities, as a signpost for psychological safety. The thematic analysis was rigorous and identified novel insights

into educator behaviours associated with indications of a safe learning environment. This article adds to the literature by drawing on the approaches of clinical educators as they attempted to invite learners into honest discussion and collaborative planning.

Our work has some limitations. We could only observe associated educator behaviours and postulate effects on psychological safety, inferred from learners' inputs or lack thereof. It is possible that learners freely expressed themselves despite feeling unsafe, although we attempted to mitigate this by focusing on occasions when learners became increasingly candid throughout the session. Further research is required to test our propositions. We did not record how long educator-learner pairs had worked together. Previous encounters would have likely influenced psychological safety within the recorded feedback episodes. In addition, participating in the study and videoing the session may have influenced participant behaviours. All participants worked at a single large health service in Australia and hence this may limit generalisability as feedback is influenced by culture and context. Other factors that may effect psychological safety, such as the individual's preferences for learning and communication, the local unit and institution involved were not investigated.[230]

6.7 Conclusions

There is increasing evidence to suggest that feedback, when characterised by an interactive learning dialogue, results in better outcomes for learners.[6,29,113,205] However, the literature is thin when it comes to explaining how a safe climate is built, to encourage learners to take productive learning risks within a feedback dialogue. Our data demonstrated than when learner and educator perspectives were shared, considered and responded to, new co-constructed ideas were fashioned. There were indications of learners 'testing the waters', with educators' responses strongly influencing subsequent learner involvement. When there were signs of psychological safety, conversations took on a 'building on' tenor, which incorporated the agendas of both parties. If participants set their sights on optimising learning through feedback, creating a psychologically safe environment to encourage learners to take risks, becomes crucial.

This in-depth observational study clarifies what psychologically safe workplace feedback conversations could look like and offers educators linguistic strategies that may foster learners' contributions to feedback 'created with them', not 'directed at them'.

6.8 Chapter 6 Closing discussion

This article included an extended excerpt from one video (see supplementary information: S6.2). It offers an illustration of how a conversation can evolve when a learner takes a 'risk for the sake of learning' and the educator responds in ways that appear to build psychological safety, illustrating each of the four key themes.

There were a number of valuable outcomes from this particular study. Psychological safety was posed as a concept that may encompass the various descriptions of similar ideas used in the feedback literature, such as 'relationship building', 'educational alliance' or 'trust'. It revealed novel insights into the strong connection between psychological safety and learner engagement in feedback. The idea of 'dipping a toe in the water' was used as an analogy to describe how a learner's perception of the degree of psychological safety seemed to influence whether they advanced or retreated from getting involved with learning and collaborating with the educator. This analysis revealed that learners are not only alert for pointers about how to improve their clinical performance, but are also vigilantly monitoring cues about the potential risk of advancing or withholding in this one-to-one encounter. As a direction for further research, it would be important to learn more about the learner's experiences of psychological safety in feedback, including their internal dialogue relating to how vulnerable or supported they feel and what they are willing to risk for the sake of learning. An approach such as video reflexive ethnography would work well in achieving this research aim.[231]

The four themes *(setting the scene; educator as ally; educator focuses on continuing improvement; and educator values interactive dialogue)* described educator approaches that appeared to cultivate psychological safety with learners and encourage them to engage in an interactive learning dialogue. This created the basis for both revising items and developing new items to present explicit and simple descriptions of observable behaviours with the potential to foster psychological safety with learners. All the refinements to the provisional instrument are outlined later in Chapter 7 (see

supplementary information Table S7.8). Here, to illustrate how this analysis led to the development of new items in the FQI, a few examples are described below.

One theme, e*ducator as ally*, described how educators showed that they saw themselves as 'on the same team' as the learners. This theme provided real-world examples of ways that educators could 'show respect', which included sharing power, actively trying to reduce the power gap by recognising their own limitations and appreciating learners' valuable contributions. These were used during the refinement process to create new items to replace this previous item in the provisional instrument:

Best interests at heart

The educator showed respect and support for the learner

The new items included:

Respect learner's autonomy

The educator showed respect for the learner's autonomy.

(This may include encouraging the learner to take a turn to lead the conversation; state their opinion or preference; make a choice; or contest the educator's comments for the purpose of learning).

Show humility and recognise own limitations

The educator conveyed the view that everyone has limitations, including themselves.

(This may include acknowledging educator's limitations e.g. routine uncertainty during clinical practice; beyond their speciality; educator's evaluation, opinion or advice contestable; made mistakes themselves while learning; or general limitations e.g. 'always more to learn' or 'a common mistake').

Appreciate learner's contributions

The educator expressed appreciation for the learner's contributions.

(This may include the learner's input into the discussion; learner's contribution to healthcare practice; learner's attributes, skills, or future potential).

The next and final paper in this thesis presents the psychometric analysis of the ratings data resulting from administration of the provisional instrument (described in Chapter 4) and the development of the Feedback Quality Instrument following refinements to the provisional instrument.

Supplementary information

Table S6.1: Coding framework for psychological safety in feedback research.

	Codes	Definition
1	• **Testing the Educator / Climate**	Learner tests educator's safe-ness (are they interested, non-judgemental, empathetic
1a	Learner testing	Learner hints or expresses mild difficulty/limitation/challenge
1b	Educator response safe	Educator demonstrates safety
1c	Educator response not safe	Educator response neutral or lack of safety
1d	Learner follow on	
2	• **Indicators of learner feeling safe**	**Learner participates, raises difficulties or limitations, gives their opinion, asks a question, offers alternate perspective or disagrees with educator**
2a	Learner participates / talks	Learner talks (use other code if more specific one fits)
	Learner self assessment	
2b (see 10e: L contests)	Learner **opinion**	Learner expresses their opinion or judgement.
2c	Learner vulnerability	Learner reveals limitation or difficulty with their own practice (incl. vulnerable self Ax)
2d	Learner seeks information	Learner asks a question or seeks more information from educator to clarify / find out more
2e	Learner laughs freely	Learner laughs (humour, not nervous laugh)
2f	Learner describes **aims**	Learner describes what they were aiming to do
2g	Learner **reflects**	Learner reflects on experiences or knowledge
2h	Learner **suggests own ideas** for improvement	Learner suggests what they could improve / how
3	• **Educator demonstrates they VALUE learner's perspective**	**Educator seeks, listens attentively, explores & responds to learner's input (participation / perspective)**
	Naming	E uses L's name
3a	Educator seeks learner's priorities	Educator finds out what learner's priorities / agenda
3b	Educator seeks **learner's perspective**	Educator seeks learner's perspective This will include: general invitation for comment / invitation for self-assessment / asking learner a question / seek response on E's comments
3c	Educator **explores /probes** learner's offering	Educator asks a <u>follow up question</u> to further explore a topic or asks a question in response to learner raising a topic
3d	Educator **summarises** (active listening)	Educator summarises learner's comments, which shows they've been listening

3e	Space	Educator leaves gaps (silence) to give learner opportunity to speak
3f	Educator's use of '**encouragers**' Desu Ne	Linguistic cues the Educator uses to convey 'I'm listening' / 'go on' / agree with learner's comments as they talk. Not interrupt / 'please continue'-ers
3g	Educator asks for feedback	Educator asks L for feedback
	E **agrees with L's evaluation** *Use other node	Use 'E validates L's perspective
4	• **Educator shows they <u>do not</u> VALUE learner's perspective**	
4a	Educator dominates	Educator does not seek learner input
4b	Educator ignores	Educator ignores topic raised by learner or changes topic
4c	Educator interrupts	Educator interrupts or talks over learner
5	• **Educator shows care for the person / respect autonomy & expertise / potential the learner**	Educator shows empathy / interest in more than learner's work performance
5a	Educator shows **interest** beyond work performance	Educators asks about learner, beyond current performance eg asks about them as a person / their wellbeing / career goals
5b	Educator shows **empathy**	Educator shows compassion for learner / recognises the learner's struggles / recognises difficulties incl 1st time etc
5c	Educator shows **care**	Educator communicates care: says something kind, considerate, supportive or encouraging (that is not empathy)
5d Respecting Autonomy	Educator offers their comments as 'for consideration', **not gospel truth**	Educator offers their comments gently and cautiously ie open to 'comeback' / 'worth thinking about'. Not 'final language'
5e Respecting Autonomy	Educator **not dogmatic** (my way)	Educator recognises that other experts do it differently / variety of ways task can be done effectively
5f Corrective fb	Educator is careful with correction	Educator is careful (might include: offers comments as an opinion / clear about what doesn't need correcting/ asks their opinion on their comments) when raising performance gaps or correcting learner.
5 Correction: Clear	Educator gives clear correction	
Correction: Mealy-mouthed	Educator gives meandering correction	Educator speaks in long, hesitant strings/ buried/ hard to decipher
5g Inclusive/ connection	Educator **includes** learner in community of practice	Educator comments or infers learner is included in 'community of HPs' or you're 'one of us'
5h Expertise	Educator recognises learner's expertise / potential	Educator makes comments that acknowledge learner's expertise / potential as a HP

5i Strategies (see 2f / 7c	Educator asks learner about the strategies they used or goals they were aiming **E asks 'What did you try?'**	Educator asks learner about the strategies they used – ie gives them credit for their efforts & developing framework
5j Mistakes OK	Educator indicates that mistakes (are) part of learning	Educator accepts learner's mistakes or limitations ('it's OK') / this is a common problem (others do this too) / focus on continual improvement / mistakes are an opportunity for learning
5k	Educator indicates the learner would not be expected to know/do that yet learning is a journey: **Learning journey**	Educator refers to the learning journey . Especially if indicates that X is beyond their expected stage of learning or that learning is a journey or it will get easier or confirm where L is on the LJ (eg this is your first one)
5l	Educator shows respect for learner's **autonomy**	Educator comments indicate the learner has autonomy / can choose eg how would you like to do this?
5m	Educator is transparent / honest	Educator is transparent about their opinions, agenda etc
6	• Educator shows a <u>lack of care</u> for the person / respect for their autonomy	
6a	Educator is **rude** [V23]	Educator makes rude or insulting comments about the learner or is verbally harsh/unkind/insensitive/ 'rough'
6b	Educator **excludes** learner from COP	Educator makes exclusionary comments
6c	Educator **overrides** learner	Educator directs or over rides learners opinion or choice
7	**• Educator ASSISTS learner**	Educator suggests ideas or solutions / offers help to assist learner or shows commitment to learner
7a	Educator explains process & agenda	Educator proposes how the session will run
7b	**E offers ideas** Educator offers help: ideas / solutions: improvement advice	Educator suggests ideas or solutions / offers help to assist learner
7c	**E scaffolds** learner's self-discovery or reflection	Educator guides them to work it out for themselves by asking questions that will lead them to the answers including helping L to make clear goals and action plans (the focus here is not on whether G & AP are made but if / how E supports L to do it)
7d	**Educator** clarifies standards (text book) **[Do it like this]**	Educator explains or clarifies standards 'what you are aiming for in your practice'
7e	Educator clarifies standards (personal practice disclosure)	Educator explains or clarifies standards by sharing what they do in their own practice (storytelling / stories of a practitioner)
7f	Educator shows their commitment to **collaboration** 'Do this together'	Educator indicates their commitment to working alongside the learner to help them
7g	Educator explains reasons	Educator explains the reasoning for their comments
7h	Educator **explains basis** / foundation	Educator explains the basis for their comments eg what they saw/ who they spoke to/ their own expertise in this area – ie reveals evidence and builds credibility
7i	Educator **seeks questions** or checks understanding	Educator seeks questions or checks understanding

8	• **Educator Vulnerability**	
8a	**Just like me**	Educator tells learner that it was the same for them as a trainee
	E apologises	
	E takes blame	E takes / shares responsibility for problem
9	• **Unity**	Educator and learner start aligning their words or behaviour and saving other's face
9a	**Educator echo's** learner	Educator uses a phrase previously used by the learner (echo it)
9b	**Learner echo's** educator	
9c	**Educator mirrors** body language of learner	Educator copies body position or movement of learner
	Learner mirrors body language of educator	Learner copies body position or movement of educator
9d	**Educator saving face for L**	Educator tries to maintain face of other
9e	**Learner preserves face** of other	Learner tries to maintain face of educator
9f	Learner and educator **laugh together**	Shared humour
10	• **Collaboration & interaction**	
10a	Building	Educator or learner use other's comments as a basis for their comments and extends it
	Collaboration / Do it together / we	
10b	Alters perspective	Educator or learner alters their own perspective after hearing other's comments
10c	Shared perspective	Educator or learner shows they have a shared understanding, goal or plans
10d	Educator **contests L**	Educator challenges or disagrees with learner's input
10e	Learner **contests** E	Learner expresses alternate opinion or disagrees with educator's input
10a	Educator **commends** learner	Educator commends learner for their attitude, effort, activities, progress or achievement
11	L compliments or thanks E	Learner commends educator
	Learner not appear to feel safe	
	Learner makes excuses or is defensive	Learner spontaneously gives reasons why they didn't do as well / excuses / I've got better Not fair
	E explore excuse	
	E does not explore excuse	
	L back tracks from raising an issue	

Supplementary information S6.2

Extended excerpt showcasing what psychological safety could look like and how educators can foster it

The following abridged excerpt offers a more comprehensive insight into an interactive learning dialogue identified within our data set. The context is a mid-attachment feedback session involving a junior doctor training to be a physician, during an attachment with a psychiatric unit:

> *Educator: It's midway through the rotation, so it's a good opportunity to…touch base and see where things are at, see how you feel things are going, and…make sure things are on track. So, from your perspective, how do you feel the…rotation is going so far?*
>
> *Learner: For me, the rotation has been very amazing. Psychiatry has always been my weakest area during my medical training, and that's one of the main reasons that I have requested to come here. [I] don't normally do a comprehensive psychiatric assessment…so I find that it is quite challenging. But, on the other side, I'm learning combining mental health along with the medical issues…I find that…very rewarding…In the management of distressed [psychotic] patients, I find that's quite challenging, like the other day with [patient's name]…I didn't realise that's how you would manage [it]…I would love to know more and improve…*
>
> *Educator: And particular things that you feel have gone well?*
>
> *Learner: I [fitted] in well with the team from the beginning,…dealing with [medial conditions in] patients..[which I am] quite used to.*
>
> *Educator: I think there are a lot of things that have gone really well…you fitted in really well with the team…your communication…respect…has been really nice…and people have liked you…Also [with the patients] you've got a really lovely manner…you are very hardworking…medically, your knowledge has been…invaluable for us…and it's very clear that you know your stuff.*
>
> *In terms of dealing with distressed patients…that is something that will come with experience…but it can...emotionally make us feel…quite uncomfortable…I've noticed you were a little bit uncomfortable a couple of times when we used the Mental Health Act... Those kind of things are very different in psychiatry from medicine…It's always, "First do no harm." So we don't lightly use it…taking someone's autonomy away but…sometimes [it is] very necessary to do.*
>
> *Learner: One more thing that I need to learn is…how to write a comprehensive report…I need [a better] structure.*
>
> *Educator: That's what I was going to say. I don't think your structure is quite clear…a lot of scratching out and filling in, so it makes for a very messy file. And that's never a good thing because people then don't read it and your plan gets lost and people don't follow up on it….That's something that's easily fixed, it's not*

a significant thing…You are very keen to learn and…always ask questions, and you are always very attentive, so I think it's just something you've not had a lot of experience [with],…you mentioned…it's not your area of specialty…I think [this is an] area that we can grow on.

Learner: …Because how I think isn't clear, I think it shows in my writing... quite different from medical [reports which I am] so used to. When patients are talking, I [am] writing what they're saying… Then I want to make a comment, and I don't know how to put it together.

Educator: Okay, so you write verbatim what's being said…

Learner: I'm thinking whether I should, next time, do a scrappy draft…and then write it.

Educator: I think the key things would be to make the management plan very clear, with the rationale. Often it pays to sit back and just listen for a minute or two and then write a summary of what's said, rather than just word for word. [V19]

The educator starts by highlighting the aim to 'make sure things are on track' and incorporating the learner's opinion, which sets the direction and tone for the interaction. The discussion is opened by the educator 'giving the learner the floor' to express their perspective on the attachment. In the video (not all included here), the learner talked for some time while the educator nodded their head, asked supplementary questions and intermittently summarised the learner's comments. This combination of curious enquiry with attentive listening skills seemed to sustain the invitation for the learner to speak and signalled that the educator valued the learner's contributions. This approach recurred throughout the conversation. The educator responded to the learner's candid comments about their difficulties and successes, by commending the learner's communication and collaboration skills, hard work and medical expertise (which the educator humbly admitted was beyond their area of expertise). Next the educator focused on one of the issues raised by the learner and offers expert insights (using inclusive 'we' language) and empathic comments about the learner's discomfort around invoking the Mental Health Act and compulsory treatment orders. Following this, the learner raised a difficulty, stating their report writing needed a better structure but they did not explicitly ask for help; this may be seen as the learner 'dipping their toe in the water'. The educator again deferred to the learner's agenda but assumed the learner just needed to focus on writing notes more neatly. They highlighted the value of clear medical notes, indicating the educator's aim was to help the learner provide quality patient care. Additionally the educator praised the learner's learning strategies ('keen to learn', 'attentive', 'always ask questions'), demonstrating they believed the learner's substandard report writing was due to their inexperience in psychiatry and were confident that the learner could easily correct this. These seem to indicate the educator has a 'learning orientation'. All these components seemed to build psychological safety. In response, the learner became increasingly candid about why their notes were messy and mentions one strategy they had thought of which could achieve a better result but would take much longer. This disclosure enabled the educator to propose a more targeted solution. Unfortunately, the educator did not

check back with the learner if they thought the educator's suggestion was useful or wanted any further clarification, to complete the communication loop. Overall, this excerpt illustrates how the key themes we have described may draw a learner into an open learning dialogue by promoting psychological safety.

Chapter 7

The Feedback Quality Instrument: A guide for health professional educators in fostering learner-centred discussions

7.1 Chapter 7 Introduction

Phase 3 focused on testing, analysing and refining the provisional instrument to produce the Feedback Quality Instrument. Phase 3 had 3 stages. In Stage 1, videos of authentic feedback interactions were collected. These were allocated in random sequence to assessors who evaluated educators' feedback practice using the provisional instrument. The ratings data were then used to rank components of quality feedback from most to least 'commonly seen'. Identifying important feedback behaviours that were rarely seen signalled that educators may benefit from additional guidance. Stage 2 concentrated on using qualitative and additional quantitative analyses to identify constructive ways to improve the provisional instrument. Qualitative analyses, described in the previous two chapters (Chapters 5 and 6), afforded deeper insights into ways that educators could foster psychological safety with learners and encourage learners to develop their evaluative judgement. This analysis also guided refinement of relevant items. The additional quantitative analysis analysed how well the data provided by the provisional instrument functioned as a measurement tool. The following chapter (Chapter 7) details this psychometric analysis of the provisional instrument, completing Stage 2, and then Stage 3 involving the refinement of the provisional instrument based on all the preceding work: the usability testing, psychometric analysis, qualitative analyses and including mapping to underpinning research and theory.

Psychometric analysis of the provisional instrument was conducted using exploratory factor analysis (EFA) and multifaceted Rasch model analysis (MFRMA). EFA is often used during the development of an instrument to distinguish clusters of closely inter-related items that reveal a smaller number of underlying concepts (factors), which summarise, or represent, those items. Ideally such factors would be represented by a number of items, to enhance accuracy. It was the results of the EFA that instigated the qualitative

analysis on psychological safety, as one cluster determined to be characterising psychological safety contained two highly correlated items, and this represented the category with the smallest number of items. Additional items were needed to describe observable educator behaviours that embodied psychological safety.

While many readers are familiar and comfortable with reading EFA methods and results, Rasch analysis remains a relatively specialised analytical approach. It is particularly valuable in instrument development for a number of reasons. An instrument that fits the Rasch model provides measures of a single underlying construct, in this case an educator's 'proficiency in orchestrating quality feedback'. Items on such an instrument would ideally target the full range of proficiency, indicated by representative observable behaviours, and the model enables them to be ranked from very simple items (those commonly done) to very hard items (those rarely done). In this way, the scale effectively acts like a 'ruler'. Rasch model analysis identifies aspects of the scale, particularly the items and rating options, that do not usefully contribute to, or may even detract from, accurate measurement of an educator's proficiency level. It enables instrument refinement by identifying whether the scale is broad enough to capture all levels of proficiency, whether items are distributed in a regular fashion across the scale, if there are gaps in the scale where additional items are required, or if there is crowding in sections of the scale where some items are not needed and can be removed.

There are two additional potential benefits that arise from the Rasch analysis. Ranking items in order of difficulty clarifies which items are 'most difficult', that is, those items that were rarely rated as 'consistently done'. This could help facilitate targeted education of clinical educators, especially as some of these behaviours may be more 'unknown' than 'difficult'. The advantage of creating a linear interval scale, which has equal intervals, from ordinal Likert-type scale data, which is ordered but with intervals of unknown size, is that it enables future research to investigate the effect of different educator behaviours on learner outcomes, by allowing comparisons between different levels of quality feedback proficiency.

Recent evolution in the field of Rasch analysis led to a decision to move from the initial analysis using a simple Rasch model to a multifaceted Rasch model for analysing data gathered using the provisional instrument. Multifaceted Rasch models can take into account additional parameters (called 'facets) that might influence the score, beyond just the educator's proficiency in single-parameter models. In this case, 'rater severity' needed to be included, as six different raters each independently rated each video. To do this required additional time to become familiar with the new Rasch skills required and then to repeat the analysis. The advantage was improved confidence in the outputs, as the application of cutting-edge modelling more accurately identified potential problems, which influenced refinements to the instrument and will underpin more sophisticated future refinements.

Refining the provisional instrument and finalising decisions on the Feedback Quality Instrument was complex. For example, both EFA and MFRM only provide indicators, not specific directions regarding 'what to do'. Recommended actions, accompanied by the rationale, were proposed by the primary researcher (CEJ) then decisions were made, taking account of multiple considerations during extensive discussions, by the core research team (CEJ, JLK, EKM, ML).

The following article will be read as a stand-alone study when published, so the research is described in two phases. Phase 1 in the article summarises work related to the development of the provisional instrument (correlating with Phase 2 in this thesis), completed and published previously; and Phase 2, the focus of this article, details the testing, analysis and refinement of the provisional instrument (correlating with Phase 3 in this thesis). To make the body of the article more accessible to a broad readership, the more detailed information, particularly regarding the EFA and MFRM analyses have been placed into supplementary information (Chapter 7 supplementary information, in this thesis).

Data analysis

The details regarding my contribution to the statistical analyses presented in the following chapter are detailed here.

The exploratory factor analysis was performed by me, using IBM SPSS Statistics for Windows, V25.0, 2017 (IBM Corp, Armonk, NY) and confirmed during a single consultation with the Statistical Consulting Centre, The University of Melbourne.

Rasch analyses were performed across two periods. A simple one-parameter Rasch model was used in the first analysis, which I performed myself after attending a three-day course (Introduction to Rasch Analysis by Associate Professor Julie Pallant) and subsequent private consultations with Associate Professor Julie Pallant. Analysis was conducted using RUMM2030 Standard edition, 2010 (RUMM Laboratory Pty Ltd, Perth, Australia). For the subsequent analyses, a multifaceted Rasch model was used. Although this is based on similar techniques, it required different software. To access this I collaborated with Peter Congdon, Manager, Assessments, Royal Australian and New Zealand College of Psychiatrists, Australia). He conducted the analysis using ConQuest Generalised Item Response Modelling Software, Version 4, 2015 (Australian Council for Educational Research, Camberwell, Victoria) and we collaborated in interpreting the results. My decisions were informed by extensive reading on multifaceted Rasch model analysis, particularly Many-faceted Rasch measurement,[232] Rasch analysis in the human sciences[233] and Applying the Rasch model: fundamental measurement in the human sciences.[234]

The Feedback Quality Instrument: A guide for health professional educators in fostering learner-centred discussions

Christina E. Johnson[1]; Jennifer L. Keating[2]; Michelle Leech[3]; P. Congdon[4]; Fiona Kent[5]; Melanie K. Farlie[5]; Elizabeth K. Molloy[6]

[1] Department of Medical Education, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne; and Monash Health, Melbourne, Victoria, Australia

[2]Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine Nursing and Health Science, Monash University, Melbourne, Victoria, Australia

[3]Faculty of Medicine, Nursing & Health Sciences, Monash University and Monash Health Melbourne, Victoria, Australia

[4]Royal Australian and New Zealand College of Psychiatrists

[5]Faculty Medicine, Nursing and Health Sciences, Melbourne, Victoria, Australia

[6] Department of Medical Education, Melbourne Medical School, Faculty of Medicine, Dentistry and Health Sciences at the University of Melbourne, Melbourne, Victoria, Australia.

Keywords: Feedback, Clinical practice, Health professions education, Educator behaviour, Instrument development, Delphi process, Rasch analysis, Exploratory factor analysis.

Acknowledgments: The authors are deeply grateful to the clinicians and students who generously participated and recorded their feedback interactions, to support this research into practical ways to improve feedback practice. We wish to thank Professor David Boud, Director, Centre for Research in Assessment and Alfred Deakin Professor, Deakin University; and Professor Debra Nestel, Professor of Simulation Education in Healthcare, Monash Institute for Health and Clinical Education, Faculty of Medicine, Nursing & Health Sciences, Monash University for their insightful comments on earlier versions of the instrument and this article.

Funding**:** none

Declaration of interest: None

7.2 Abstract

Context: Face-to-face feedback plays an important role in health professionals' workplace learning. The literature describes guiding principles regarding learner-centred feedback but it is not clear how to enact these. We aimed to create a Feedback Quality Instrument (FQI), underpinned by a social constructivist perspective, to assist educators in collaborating with learners to promote learner-centred feedback interactions. In earlier research, we developed a set of observable educator behaviours designed to foster beneficial learner outcomes, supported by published research and expert consensus. This research focused on testing and refining this provisional instrument, to create the FQI ready-to-use.

Methods: We used a mixed methods design. We collected videos of authentic face-to-face feedback discussions in clinical settings. Quantitative and qualitative analyses of the video data were used to refine the provisional instrument. Raters administered the provisional instrument to systematically analyse educators' feedback practice seen in the videos. This enabled usability testing and resulted in ratings data for psychometric analysis involving multifaceted Rasch model analysis and exploratory factor analysis. Parallel qualitative research of the video transcripts focused on two under-researched areas, psychological safety and evaluative judgement, to provide practical insights for item refinement. The provisional instrument was revised, using an iterative process, incorporating findings from usability testing, psychometric testing and parallel qualitative research and foundational research.

Results: Thirty-six videos involved diverse health professionals across medicine, nursing and physiotherapy. Administering the provisional instrument generated 174 data sets. Following refinements, the FQI contains 25 items, clustered into five domains characterising core concepts underpinning quality feedback: *set the scene*, *analyse performance*, *plan improvement*, *foster learner agency*, and *foster psychological safety*.

Conclusions: The FQI describes practical, empirically-informed ways for educators to foster quality, learner-centred feedback discussions. The explicit descriptions offer

guidance for educators and provide a foundation for systematic analysis of the influence of educator behaviours on learner outcomes.

7.3 Background

In the health professions, face-to-face feedback plays a key role in workplace learning and can have a powerful impact on performance.[235] Common feedback approaches include more scheduled, comprehensive performance discussions, for example workplace-based assessments or end-of-attachment appraisals; or more brief impromptu comments or tips offered while delivering clinical care (often called 'feedback on the run'). Recent feedback literature, underpinned by social constructivism, supports learner-centred feedback conversations in which learners actively participate, to acquire knowledge they can use to enhance subsequent performance.[19,26,199,208] A performance discussion with an educator offers opportunities for a learner to advance their understanding of the key characteristics of the target clinical performance ('where am I aiming for?'), how their own performance compares to this ('where am I now?'), and work out what they can do to improve ('how can I get closer?').[1,6,18,113] When learners and educators collaborate through an interactive dialogue, together they can generate new performance insights and strategies for improvement, individually tailored for the learner.[20,43]

However, the literature does not provide clear guidance on how to apply these principles in practice; that is, what can educators and learners do to enact learner-centred feedback? Studies have identified a gap between recommended and observed practices. Frequently, educators dominate feedback episodes and learners play a passive role.[57,114,115] Learners report that often they do not find educators' comments relevant, and struggle to understand or apply the information.[11,13,119,236] Educators typically undertake minimal training in feedback (when contrasted with the rigorous development of clinical skills) and report a lack confidence in their feedback skills.[12,36,38,123,237] It may be that, in the absence of alternative strategies, educators are simply repeating feedback rituals they experienced as students or using formulaic assessment rubrics, which are not designed with an *interactive process* in mind. Hence there is a need for new schemas that are structured to promote educator and learner collaboration during feedback interactions.[27,238]

A number of feedback models have been described in health professions education literature.[143,154,167] These provide useful insights to assist educators' feedback practice. Some were designed for specific contexts such as formal discussions regarding written performance assessments,[168] experiential communication skills training,[154] or debriefing in simulation-based education.[143] Many of these guiding models were developed based on expert opinion, focused literature reviews or theoretical perspectives (or combinations of these). A few have reported modifications based on testing, such as inter-rater reliability or usability testing.[143,168,206]

Our research program is focused on assisting educators to facilitate high quality, learner-centred, feedback interactions. It is based on a social constructivist paradigm, in which learners actively build and refine their mental schemas during interactions with other people at work.[20] We have focused on the educator, as 'one partner in the dance', because educators typically have a major influence on feedback interactions and have a responsibility to promote rich learning opportunities.[27,239] Our goal is to create an instrument, the Feedback Quality Instrument, to guide educators in high quality learner-centred feedback. This instrument describes observable educator behaviours to promote collaboration with learners during feedback to enhance learner outcomes. By providing exemplars of helpful behaviours, this instrument could contribute to clarifying 'what quality feedback looks like' and assist educators to develop their feedback practice. In addition, a comprehensive set of observable educator behaviours could enable systematic analysis of which feedback components, or combinations, have the greatest beneficial impact on learner outcomes.

The development of the Feedback Quality Instrument is described in two phases. Phase 1 involved creating a provisional instrument (previously published)[52] and Phase 2, the focus of this article, involved testing and refining it.[240,241] In summary, Phase 1 contained the following three stages (see Figure 7.1):

Stage 1: Clarifying the construct, that is what constituents needed to be included in the instrument, using an extensive review of the literature to identify discrete elements of an educator's role considered to influence learner outcomes, supported by empirical

information. The review identified over 170 relevant articles across involved health professions education, education, business and psychology literature and included analyses of feedback observations, forms, surveys and interviews; feedback models; systematic reviews; consensus documents; and educational and psychological theories;

Stage 2: Generating initial items, using an iterative deductive process to convert the elements into representative observable educator behaviours;

Stage 3: Expert refinement of the initial item set, using three rounds of a Delphi process that led to consensus on a provisional instrument with content validity.

Hence Phase 1 resulted in a provisional instrument, incorporating a set of observable educator behaviours designed to support learners' engagement, motivation and capacity to improve during feedback discussions in clinical practice. The provisional instrument is reproduced in Figure 7.2.[52]

The purpose of this current research, Phase 2, was to test and refine the provisional instrument, and present the Feedback Quality Instrument, validated and ready for use in clinical practice. For Phase 2, our research question was:

In what ways can the provisional instrument be refined, based on usability testing, psychometric analysis and parallel qualitative analyses of video data of authentic feedback interactions, to produce the Feedback Quality Instrument?

Figure 7.1: Development of the Feedback Quality Instrument: Completed Phase 1, Stages 1-3 to create a provisional feedback instrument. [52]

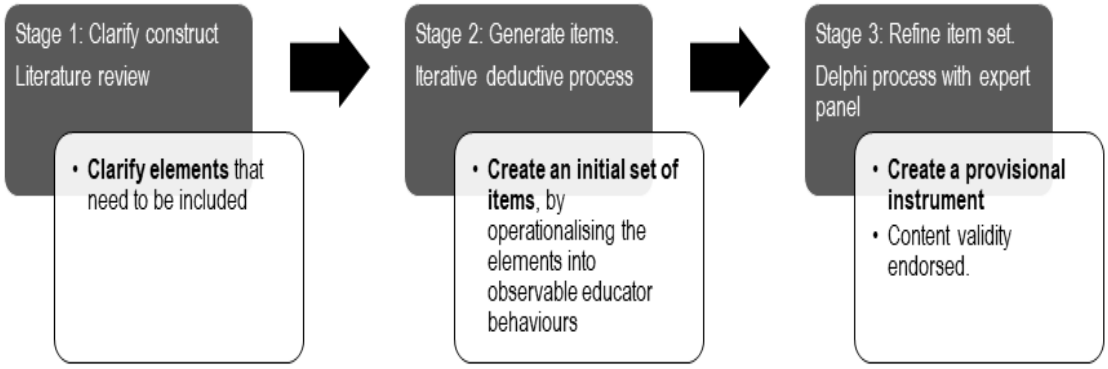


Figure 7.2: Set of items constituting a provisional feedback instrument.

Recommended educator behaviours in quality face-to-face feedback discussions in clinical practice

Item 1: *Based on observed performance*

The educator's comments were based on observed performance.

Item 2: *Timely feedback*

The educator offered to discuss the performance as soon as practicable.

Item 3: *Feedback purpose clear*

The educator explained that the purpose of feedback is to help the learner improve their performance.

Item 4: *Establish a non-judgmental atmosphere: 'here to help'*

The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise.

Item 5: *Clarify feedback process, so learner knows what to expect*

The educator described the intended process for the feedback discussion

Item 6: *Encourage dialogue*

The educator encouraged the learner to engage in interactive discussions

Item 7: *Seek learner's priorities*

The educator asked the learner about their learning priorities for the observation and feedback discussion, and responded to them

Item 8: *Encourage learner to 'work it out for themselves'*

The educator encouraged the learner to consider the issues and possible solutions during the feedback.

Item 9: *Encourage learner to focus on learning, rather than trying to cover up limitations*

The educator encouraged the learner to discuss difficulties and ask questions regarding the performance so the educator could help the learner to develop solutions

Item 10: *Acknowledge learner's emotional response*

The educator acknowledged and responded appropriately to emotions expressed by the learner

Item 11: *'Best interests at heart'*

The educator showed respect and support for the learner

Item 12: *Clarify the value of self-assessment*

The educator asked what the learner understood about the benefits of self-assessment and helped clarify

Item 13: *Learner self-assessment*

The educator asked the learner to identify key similarities and differences between the learner's performance and the target performance

Item 14: *Target performance and reasoning clear*

The educator clarified with the learner key features of the target performance and explained the reasoning

Item 15: *Educator assessment, including clear performance gap*

The educator clarified with the learner similarities and differences between the learner's performance and the target performance

Item 16: *Educator comments on a few, important issues*

The educator's comments focused on key issues for improving the performance

Item 17: *Specific instance ('what happened')*

First the educator described, using neutral language, what the learner did (action, decision or behaviour), and the consequences

Item 18: Educator's perspective clear ('why it matters')

The educator clearly explained their perspective on the learner's actions, including the reason for their concern

Item 19*: Educator explores learner's perspective ('why' the learner acted as they did)*

The educator explored the learner's perspective and reasoning to reveal the basis for the learner's actions (e.g. what was the learner trying to do and options considered/ difficulties encountered).

Item 20: *Focus on actions, not the person ('did' not 'is')*

The educator's comments were focused on the learner's actions not personal characteristics

Item 21: *Select learning priorities: most useful (important and relevant) for the learner*

The educator helped the learner to select a couple of key aspects of the performance to improve

Item 22: *Develop the action plan: how to do it!*

The educator helped the learner to work how they could improve their performance and specify the practical steps to achieve it.

Item 23: *Check the learner understands the plans*

The educator checked if the learner understood their learning goals and action plan, by asking them to summarise it in their own words.

Item 24*: Checks the learner understands the rationale: 'why it's better'*

The educator checked if the learner understood the rationale for their learning goals and action plan.

Item 25: *Plan opportunities to review the impact of the feedback*

The educator discussed with the learner possible subsequent opportunities for the learner to review their progress.

7.4 Methods

Research overview

This research used a multi-phased mixed methods design. Phase 1 developed a set of 25 items, representing a provisional feedback quality instrument, briefly summarised above and described in more detail elsewhere. [52] This research describes Phase 2, which focused on testing and refining the provisional instrument using videos of scheduled feedback discussions in clinical practice, to produce the Feedback Quality Instrument. The three stages in Phase 2 included:

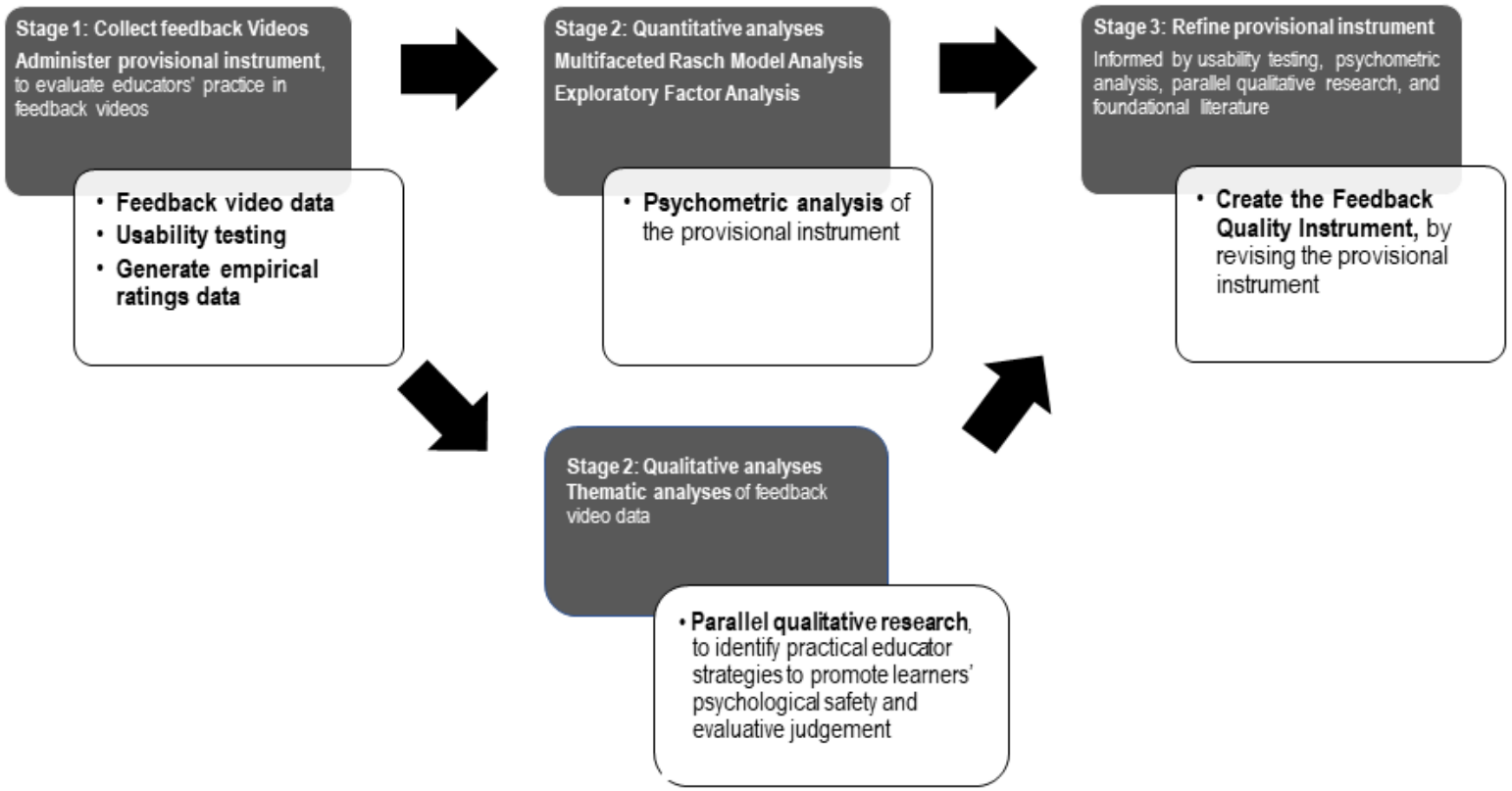
Stage 1: Collecting videos of feedback discussions in routine clinical practice. Then using the provisional instrument to systematically evaluate educators' practice seen in the feedback videos, which enabled usability testing and provided item ratings for psychometric analysis;

Stage 2: Using quantitative and qualitative analyses of the video data to refine the provisional instrument. Psychometric testing of the item ratings data was conducted using Multifaceted Rasch Model (MFRM) analysis and exploratory factor analysis (EFA). Parallel qualitative analyses of the video transcripts, reported in detail elsewhere, investigated two important but under-researched aspects of feedback, evaluative judgement[172] and psychological safety,[242] to enhance the provisional instrument. In particular, additional items were created for one instrument domain, *foster psychological safety*, as it was considered to be inadequately characterised following EFA analysis and a review of the latest literature did not reveal information on specific ways that educators could collaborate with learners to nurture psychological safety.

Stage 3: Refining the provisional instrument, informed by the usability testing, psychometric testing, parallel qualitative research and underpinning research and theory (see Figure 7.3).

Ethics approval was obtained from the health service (Reference 15233L) and the university human research ethics committees (Reference 2015001338).

Figure 7.3: Development of the Feedback Quality Instrument: Phase 2: Testing, analysis and refinement of the provisional instrument to produce The Feedback Quality Instrument.



Stage 1: Collecting feedback videos and administering the provisional instrument

Collection of feedback videos

Videos of authentic scheduled feedback sessions were collected. To recruit participants for the feedback videos, first a diverse range of educators (supervising clinicians) across medicine, nursing and allied health in a major metropolitan teaching hospital network in Australia were invited to participate. When an educator consented, learners (students or clinicians) working with the educator at the time were invited to participate by the research team. Once both members in an educator-learner pair consented, they arranged to video themselves during the next face-to-face feedback session scheduled to discuss the learner's performance in routine clinical practice. This methodology has been described in more detail previously.[57]

Administering the provisional instrument to evaluate educators' feedback practice

Raters administered the provisional instrument and compared educator behaviours seen in each feedback video with recommended educator behaviours (See Figure 7.2 for the provisional instrument). Each item was rated as 0 = not seen, 1= done somewhat, or 2 = done consistently. A pilot was conducted within the study to resolve preliminary problems using the instrument. This resulted in removal of *Item 2: The educator offered to discuss the performance as soon as practicable*, as this occurred before, not during, a feedback interaction. Subsequently all raters independently analysed all videos, which were presented in a random order devised using an online random number generator. Administration of the provisional instrument generated i) empirical item ratings data, subsequently used for psychometric analysis, and ii) usability analysis. (For more details regarding the raters and the pilot, see supplementary information: Section S7.1).

Usability analysis of the provisional instrument

While administering the provisional instrument, the rating team recorded comments regarding the usability of the instrument, items and rating scale, including both individual contemporaneous written comments during video analysis and two

scheduled team telephone discussions, which were recorded.[241] (For more details, see supplementary information: Section S7.2).

Stage 2a: Quantitative analysis of feedback video data: psychometric analysis of the provisional instrument using video ratings data

To investigate the psychometric properties of the provisional instrument, the ratings data were used to conduct 1) multifaceted Rasch model analysis and 2) exploratory factor analysis.

Multifaceted Rasch model analysis (MFRMA)

The multifaceted Rasch model analysis examined how well the provisional instrument functioned as a measurement scale for estimating educators' feedback proficiency, by analysing how closely the observed item ratings matched those expected by the model. The multifaceted Rasch model took account of the different parts of the measurement system, including items, raters or rating scale categories (each called 'facets').[234] As the aim was to refine the provisional instrument, the analysis was primarily used to highlight items, raters or rating categories that showed substantial 'misfit' to the model, suggesting they may not usefully contribute, or may even degrade, the instrument's performance as a measurement system, and may need modifying. The analysis reported a 'person separation reliability' level that indicated how well the instrument discriminated between educators with different proficiency levels. MFRMA enables the creation of a linear interval scale, rather like 'a feedback proficiency ruler', based on the Likert ratings data from the instrument. This was displayed on a 'variable map' that showed the spread of items (easy to difficult), participants (low to high proficiency) and raters (lenient to severe) on the same linear scale, which enabled comparisons between them. (For more details on the MFRMA methods, see supplementary information: Section S7.3).

Exploratory factor analysis (EFA)

EFA is a common technique used to explore the characteristics of an instrument and guide its development,[243-245] often in addition to Rasch analysis.[246,247] The exploratory

factor analysis, using principal components analysis and direct oblimin rotation was conducted to identify clusters of closely inter-related items representing 'factors', or core concepts, underlying 'quality feedback proficiency'.[243,248] (For a comprehensive description of the EFA methods, see supplementary information: Section S7.4).

Stage 2b: Qualitative analysis of feedback video data

Qualitative analyses were conducted using thematic analysis of the video transcripts focusing on two particular aspects of feedback: psychological safety[242] and evaluative judgement,[172] described in previous publications. There is increasing interest concerning these important aspirations in quality feedback in the feedback literature but little practical guidance how educators can collaborate with learners to promote them.

Psychological safety was defined by Edmondson as "a shared belief that the team is safe for interpersonal risk taking", which creates "a sense of confidence that the team will not embarrass, reject or punish someone …due to mutual respect and trust".[193(p354)] Similar concepts discussed in the literature include 'trust',[29] the 'educator-learner relationship',[3,167] the 'educational alliance'[55,99] and creating a 'safe container'.[196,197] When learners participate in learning conversations, they may expose their limitations by raising performance difficulties, explaining their reasoning or asking questions, which risks their professional reputation. At times learners choose to take this risk, in the hope of enhancing their skills and achieving their career goals. Hence it seems likely that learners' sense of psychological safety will influence their level of involvement and vulnerability during feedback discussions.

Evaluative judgement was defined by Tai *et al* as "the capability to make decisions about the quality of work of self and others".[204] Knowing 'what good work looks like' is a key skill underpinning life-long learning, as tacit standards need to be understood and applied in daily work.[26,203] Feedback interactions provide valuable opportunities for a learner to develop their evaluative judgement by analysing their performance in comparison with the desired performance. Educators can assist by encouraging learners' self-assessment, clarifying key features of the desired performance and confirming the

learner's evaluation or explaining an alternative view, to help calibrate the learner's judgement.

Due to the limited research available to assist us in revising the provisional feedback instrument, we conducted thematic analysis of the feedback video transcripts to identify how educators in our study had promoted learners' psychological safety and evaluative judgement during the feedback sessions.[127] These thematic analyses enabled the refinement of the provisional instrument by clarifying empirically informed, practical approaches for educators to support learner-centred quality feedback interactions.

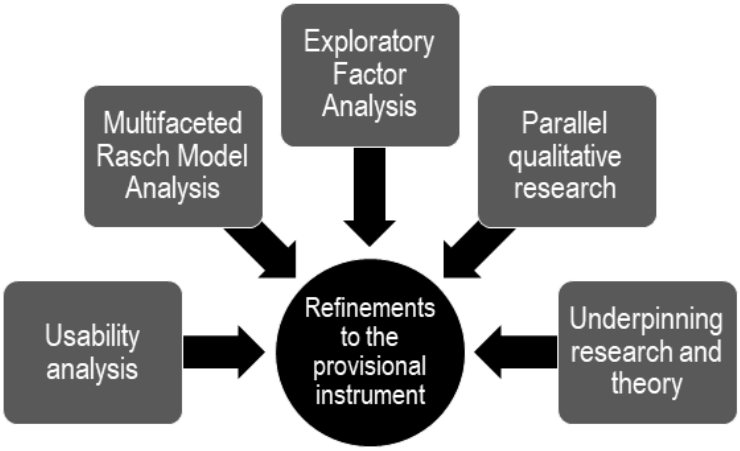
Stage 3: Refinement of the provisional instrument

The instrument and individual items were modified to better achieve the desirable criteria, previously established, that a) the instrument overall should achieve a comprehensive yet parsimonious set of items, that is, just enough items to sufficiently cover important discrete elements of an educator's role in a quality feedback discussion across the full range of feedback proficiency; b) individual items should describe an observable educator behaviour, target a single distinct attribute, be unambiguous (phrasing clear and simple, so the meaning is easily and consistently understood without further explanation), be relevant and generally applicable to a verbal face-to-face feedback interaction and make sense with each rating category; c) the rating category options should be just sufficient to cover likely possibilities, and the phrasing of the rating categories should be consistent, clear and simple.

Revisions to the provisional instrument were informed by 1) usability analysis, 2) psychometric analysis involving multifaceted Rasch model and exploratory factor analysis, 3) the parallel qualitative studies on psychological safety and evaluative judgement and 4) key theoretical principles that support learner-centred feedback, particularly relating to learning, motivation, psychological safety, evaluative judgement, and performance improvement (see Figure 7.4). Modifications to items and the instrument overall were made using an iterative process (inductive and deductive) involving multiple rounds of revision and review based on all relevant considerations by

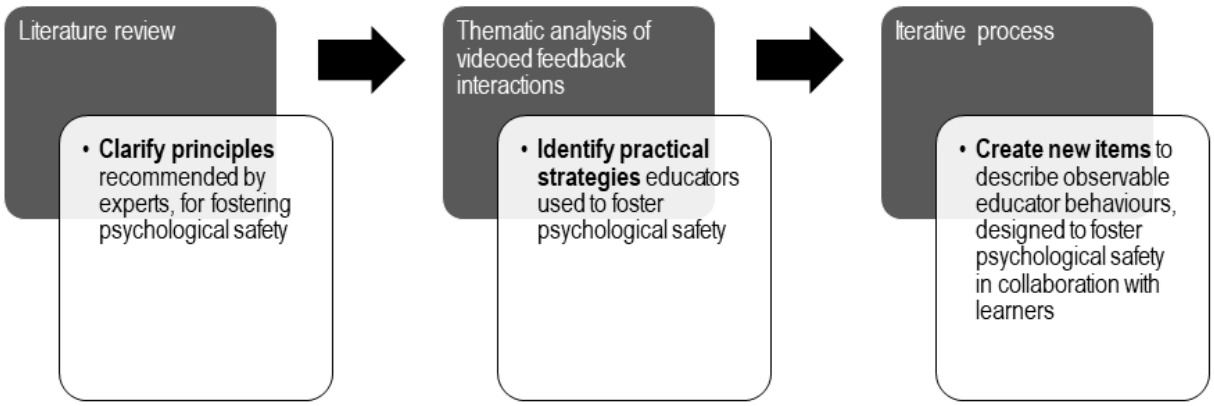
a subgroup (CEJ, JLK, EKM), in consultation with the research team and key experts from our previous Delphi panel.

Figure 7.4: The multiple inputs that informed refinements to the provisional instrument, to create The Feedback Quality Instrument.



In particular, the EFA revealed factors, involving clusters of items, that indicated core concepts within quality feedback. During the instrument revision process, items were organised accordingly, to create domains in the Feedback Quality Instrument. If a factor was considered to be insufficiently characterised by those items, this triggered a process to create supplementary items. This decision was based on 1) the number of items. It is recommended a factor contain at least three items (although two items may comprise a factor if they are strongly inter-related with each other and relatively unrelated to other items)[245] and typically, complex concepts necessitate several items to elucidate and operationalise them[243]; and 2) a further review of relevant theory and research published in the literature, to identify relevant elements. Consequently, as explained in the results, the findings from the psychological safety study were used to create additional items for one domain related to psychological safety. The findings from this study were operationalised into new items, in accordance with desirable item criteria described above, and using the same iterative process. (see Figure 7.5)

Figure 7.5: Process used to develop additional items for one domain, related to psychological safety, in the Feedback Quality Instrument.



7.5 Results

Administering the provisional instrument

Feedback videos and health professional participants

Thirty-six videos of scheduled feedback discussions during routine clinical practice were collected, involving educator-learner pairs across different health professions and specialities, experience levels and gender. In particular, there were 34 educators including 26 medical from every major speciality, 4 nursing and 4 physiotherapy health professionals. (For more details on the participants, see supplementary information: Section S7.5).

Using the provisional instrument to evaluate educators' feedback practice

Each video was analysed by four to six raters, as unexpected time constraints prevented two researchers from analysing all of the videos (1 rater analysed 21/36 (58%) and 1 rater analysed 10/36 (28%)). This yielded 174 sets of ratings data. Missing data were uncommon (0.2% ratings missing). (For more details on item ratings frequency data, see supplementary information: Section S7.5). Additional information including descriptive statistics of educators' behaviours has been described elsewhere.[57]

Usability analysis of the provisional instrument

Raters reported issues related to items 1, 7, 11, 12, 13, 17, 18, 19, such as overlapping items, ambiguous phrasing, restricted applicability or difficulty applying all the rating options, so these items were flagged for review. (For more details on usability analysis, see supplementary information: Section S7.5).

Multifaceted Rasch Model analysis

Item, rater and rating category analysis, and person separation reliability

In the MFRMA, items 5, 6, 8, 14, 15, 16 and 23 demonstrated misfit, so all these items were flagged for review, with a particular focus on items 5, 6, 14 and 23, which demonstrated misfit in the sensitivity analysis designed to isolate problems due to items themselves, especially Item 5 that demonstrated more serious misfit.

Rater severity across the different raters was fairly similar except for Rater 2, whose ratings were more severe, indicated by severe misfit. Rater severity may be modified with training but consistency in rater severity is more important and MFRMA adjusts educator proficiency scores to take account of rater severity. Rating category 1 (1= done somewhat) showed misfit, so potential reasons for this were investigated. (For more details on item, rater severity, and rating category fit, see supplementary information: Section S7.6).

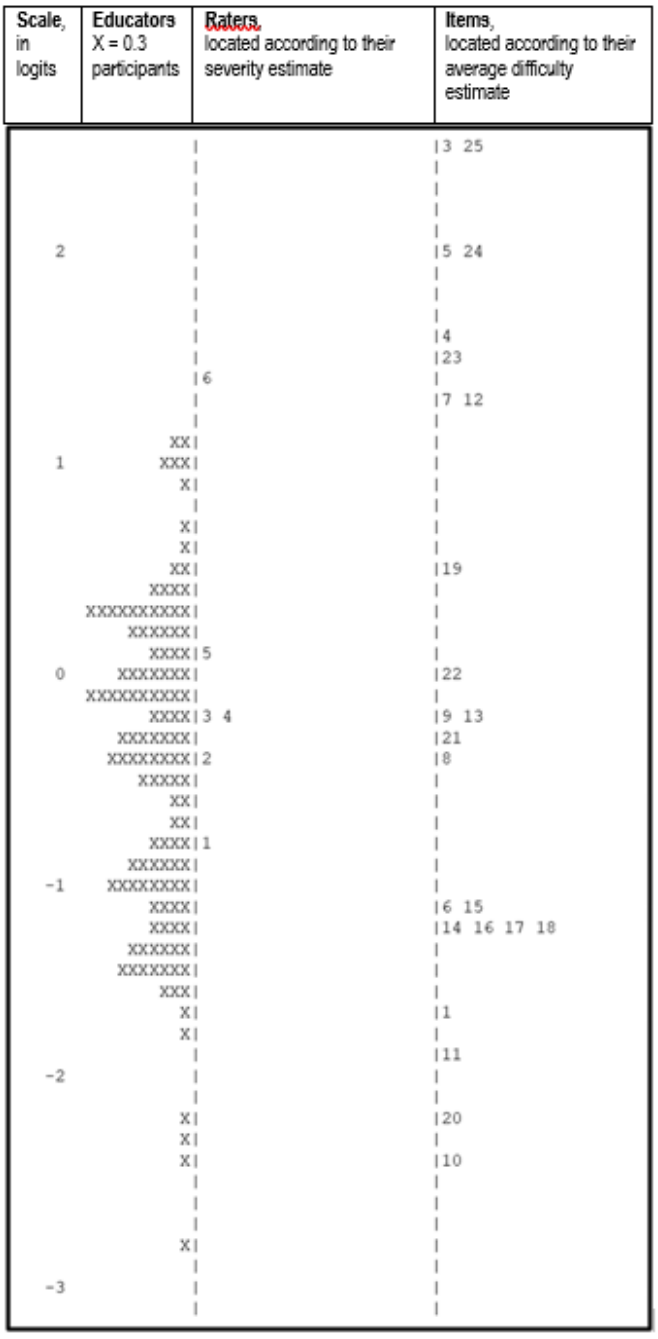
The person separation reliability was 0.95, which indicated the provisional instrument with multiple raters, could differentiate at least 4 levels of feedback proficiency amongst the educators.

Variable map

The variable map is presented in Figure 7.6. From left to right, variable map displays the linear interval scale (a 'feedback proficiency ruler', which uses a 'logit' as the unit of measurement) and distribution of clinical educator proficiency, rater severity and item difficulty on the same scale. The scale is set with the mean educator feedback

proficiency estimate at zero logits. In particular, it can be seen that items and participants are reasonably distributed across the feedback proficiency range. (For more details on the variable map, see supplementary information: Section S7.6).

Figure 7.6: Variable map showing clinical educator proficiency, rater severity and item difficulty on the same interval scale.



Footnote: Educators are shown as X = 0.3 to provide a slight distribution incorporating each participant's estimate and standard error.

Exploratory factor analysis

Exploratory factor analysis revealed five concepts (factors) that constituted 'quality feedback', represented by closely related item clusters. Four factors had multiple items that were strongly inter-related and theoretically aligned, which were named: *set the scene, analyse performance, plan improvement* and *foster learner agency*. The fifth factor only had two items but these were strongly inter-related and theoretically aligned, and it was named *foster psychological safety*. Items 5, 7 and 25 did not cluster strongly in any factor, suggesting potential problems, so these were flagged for review. (For more details of the EFA results, see supplementary information: Section S7.7).

Refinement of the provisional instrument

Multiple refinements were made to the provisional instrument, based on the results from this multi-staged analysis (see Figure 7.4, and Table 7.1 for specific outcomes, typical reasons and potential actions arising from the usability and psychometric analysis). The variable map from the MRFMA showed the spread of items across the range of feedback proficiency was acceptable with no substantial gaps, redundancy, ceiling or floor effects. The EFA identified item clusters, representing core concepts underlying quality feedback, so items in the instrument were regrouped accordingly. This provided a way to clarify the major domains and make it easier for users to understand the core concepts constituting quality feedback, instead of a large number of separate items.

From the EFA, two items (items 10 and 11) constituted a fifth factor, *foster psychological safety.* It was decided that these items alone did not adequately characterise this important concept, so a process was initiated to create additional items. These new items, which described observable educator behaviours designed to foster psychological safety in collaboration with learners, were created by operationalising the findings from the parallel qualitative study into psychological safety and the related principles identified in the literature. Item development was performed by a subgroup (CEJ, JLK, EKM) using an iterative process, combining inductive and deductive reasoning, during multiple rounds of revision and review.

In addition, the findings from the qualitative analyses into evaluative judgement and psychological safety contributed to revising relevant items (for more details on the qualitative studies' findings, see supplementary information: Section S7.8).

Individual item modifications, based on inputs from all analyses, involved merging overlapping items, improving the phrasing of items (common revisions included making the item phrasing clearer, more generally applicable, and make sense with each rating category) and adding succinct additional information to clarify further, if required. Details of the item refinements are outlined in detail in supplementary information Table S7.8. The rating scale was revised to make the phrasing more consistent across rating categories. Subsequently, the instrument rating was: *Across the feedback session, how consistently did the educator do this? 0 = not done; 1 = done sometimes; 2 = done consistently.* For once off items, for example FQI item 1, if the educator demonstrated the behaviour as described in the item, this should be rated as 2 = done consistently.

Table 7.1: Analysis outcomes, typical reasons for those outcomes and subsequent potential actions to refine the provisional instrument, arising from usability analysis, exploratory factor analysis and multifaceted Rasch model analysis.

	Analysis outcomes	Typical reasons	Potential actions
Usability analysis	**•Identify instrument problems**	• Not easy to use e.g. too many individual items or insufficient / complex instructions	• Find a way to simplify instrument administration e.g. group related items • Offer clear, useful and succinct instructions
		• Item gap	• Create new items to address gap
	• Identify item problems	• Items overlap	• Merge items
		• Item not generally applicable in a feedback session	• Remove or rephrase so item is generally applicable
		• Item phrasing: vague or not-observable description of behaviour	• Remove or rephrase so item clearly and simply describes observable behaviour
	• Identify rating category problems	• Too many rating categories, so hard to differentiate	• Reduce the number of rating categories
		• Rating category phrasing vague	• Rephrase rating category description so it is consistent, clear and simple
		• Middle rating category not applicable in some items	• Rephrase item so all rating categories are applicable
Exploratory factor analysis	**• Identify factors (core concepts) underlying quality feedback, represented by item clusters**	• Items in clusters are closely aligned i.e. all attributes of one concept	• Group items into instrument domains, and name accordingly
	• Determine if each factor is adequately characterised, with sufficient items strongly aligned with it (3 items minimum, typically)	• Insufficient items (e.g. only 2 items that strongly align)	• Create new items to describe observable behaviours that reflect that concept

	Analysis outcomes	Typical reasons	Potential actions
	• Identify items that do not align strongly with one cluster	• Item alignment split between 2 clusters (e.g. due to item phrasing or context)	• Remove or revise item, to align with one cluster
		• Item does not strongly align with any cluster (e.g. due to item phrasing problems; item behaviour not influential; or insufficient data)	• Remove or revise item, to align with one cluster
Multifaceted Rasch model analysis	**• Identify misfit shown by items, raters or rating category, which may distort the measurement system**	• Lack of consistent interpretation of item and application of rating category, due to: - item phrasing problems, so interpretation is variable - rating category problems, so application is variable • Insufficient data (if behaviour or rating category rarely employed)	Enhance consistency by • Removing or revising items and rating categories, according to desirable criteria • Using instrument manual and rater training
	• Determine spread of items across range of 'feedback proficiency' (illustrated on the variable map)	• Span with no items (gap)	• Create new items to address gap
		• Span with too many items (redundant items)	• Remove items to reduce redundancy

7.6 The Feedback Quality Instrument

On completion of this multi-phased mixed methods research process, incorporating empirical insights from the literature, usability analysis, psychometric analysis and parallel qualitative studies into psychological safety and evaluative judgement, the Feedback Quality Instrument, ready for use, is presented in (see Figure 7.7 and 7.8).

Figure 7.7: The Feedback Quality Instrument

The Feedback Quality Instrument

Item rating scale: Across the feedback session, how consistently did the educator do this?
0 = not done 1 = done sometimes 2 = done consistently

Set the scene

Item 1: Clarify purpose

The educator explained that the purpose of feedback is to help the learner improve their performance. .

Item 2: Discuss session plan, so learner knows what to expect

The educator discussed expectations for the feedback session, taking the learner's perspective into account.

(This may cover the session outline; time available; values of interactive dialogue and collaboration).

Item 3: Seek learner's priorities

The educator explored what the learner most wanted to raise in the feedback session.

Item 4: Expect mistakes and regard them as learning opportunities

The educator conveyed the view that mistakes or omissions are expected while developing skills and are opportunities to continue refining skills.

Analyse performance

Item 5: Learner self-assessment

The educator encouraged the learner to identify key similarities and differences between the learner's performance and the target performance.

Item 6: Clarify target performance and reasoning

The educator clarified with the learner key features of the target performance and explained the reasoning.

Item 7: Educator's performance analysis

The educator clarified with the learner similarities and differences between the learner's performance and the target performance.

(Consider focusing on the 'development zone' i.e. around their current position on the learning curve).

Item 8: Prioritise comments

The educator's comments focused on a few, key points for improving performance.

(This may focus on improving the current task or broader learning strategies).

Item 9: Educator's perspective clear

The educator's comments (information or opinion), accompanied by the reasoning, were clearly explained.

Item 10: Specific instance

The educator linked their comments to a neutral description of what the learner did (action, decision, behaviour) and the consequences.

Item 11: Focus on actions, not the person ('did' not 'is')

The educator's comments were focused on the learner's actions, not personal characteristics.

Plan improvement

Item 12: Select learning priorities: most useful (important and relevant) for the learner

The educator helped the learner to select a few, key aspects of the performance to improve.

(Consider learner's request; key mistake or omission or 'next steps').

Item 13: Develop the action plan

The educator helped the learner to develop specific and practical plans to improve their performance.

(The plans may relate to the current task and/or broader learning strategies).

Item 14: Check plans and rationale understood

The educator checked if the learner understood their learning goals and action plan, accompanied by the rationale, by asking them to summarise it in their own words.

Item 15: *Plan to review progress*

The educator discussed with the learner possible opportunities for the learner to review their progress.

Foster learner agency

Item 16: Promote interactive dialogue

The educator encouraged discussions in which both perspectives were shared and thoughtfully responded to.

(This may include agreeing, seeking more information, checking information, asking questions, building on an idea or contesting an idea with the rationale).

Item 17: Promote the value of self-assessment

The educator promoted the benefits of self-assessment in discussion with the learner.

(Self-assessment provides opportunities for a learner to describe key features of the target performance, analyse their own performance in comparison to the target performance and raise learning needs; this involves practising valuable learning skills such as reflection, evaluative judgement and self-regulated learning).

Item 18: Encourage learner to focus on learning, rather than trying to cover up limitations

The educator encouraged the learner to discuss difficulties and ask questions regarding the performance so the educator could help the learner to develop solutions.

Item 19: Support learner to 'work it out for themselves'

The educator encouraged the learner to consider the issues and possible solutions during the feedback discussion.

Foster psychological safety

Item 20: Value learner's perspective

The educator explored the learner's perspective and reasoning, and demonstrated attentive listening.

Item 21: Appreciate learner's contributions

The educator expressed appreciation for the learner's contributions.

(This may include the learner's input into the discussion; learner's contribution to healthcare practice; learner's attributes, skills or future potential).

Item 22: Respect learner's autonomy

The educator showed respect for the learner's autonomy.

(This may include encouraging the learner to take a turn to lead the conversation; state their opinion or preference; make a choice; or contest the educator's comments for the purpose of learning).

Item 23: Show compassion

The educator expressed compassion for difficulties experience during training, raised by the learner.

Item 24: Attend to learner's emotions

The educator responded appropriately to emotions expressed by the learner.

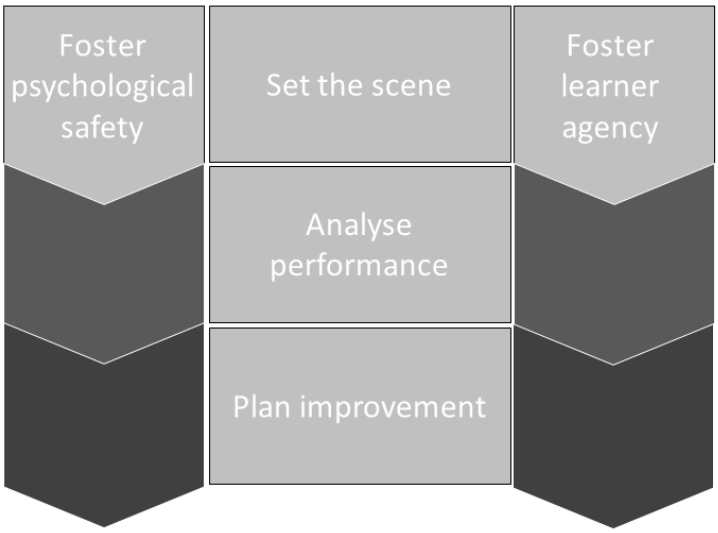
Item 25: Show humility and recognise own limitations

The educator conveyed the view that everyone has limitations, including themselves.

(This may include acknowledging educator's limitations e.g. routine uncertainty during clinical practice; beyond their speciality; educator's evaluation, opinion or advice contestable; made mistakes themselves while learning; or general limitations e.g. 'always more to learn' or 'a common mistake').

Note: items are numbered so they can be easily referred to, not to indicate a 'correct order'.

Figure 7.8: Schematic diagram showing the five domains, representing core concepts underpinning high quality feedback, within the Feedback Quality Instrument.



7.7 Discussion

In this research, we described the analysis and refinement of a provisional feedback instrument, developed earlier,[52] to create the Feedback Quality Instrument (FQI) (see Figures 7.7 and 7.8). To our knowledge, no other feedback instrument designed for clinical practice has undergone such a rigorous development process (see Figures 7.1 and 7.3). The FQI clarifies how educators can work together with learners to promote high quality learner-centred feedback discussions in clinical practice. The items describe educator behaviours designed to engage learners in an interactive learning dialogue. This moves beyond tips focused on making educators' input useful (e.g. timely, relevant, specific), to supporting learners to reveal difficulties, ask questions and refine ideas, so learners can enhance their understanding of their work, the required standards and instigate improvements. By attempting to explicitly characterise the educator's role, we hope to ignite debate and research that leads to continuing refinements. We recognise that every feedback interaction needs to be customised, so the sequence or emphasis

will vary depending on the individuals and the specific context. Additionally, it is important to enhance the capacities of both educators and learners to effectively contribute to these conversations. We have chosen to focus on investigating the educator's role in promoting beneficial learner outcomes and we recommend that readers consider complimentary work exploring ways to optimise the learner's role in feedback [26,28,181,204,249]

The FQI contains five domains, three that occur somewhat sequentially, *set the scene, analyse performance, plan improvement,* and two that continue throughout the interaction, *foster learner agency* and *foster learner agency* (see Figures 7.7 and 7.8). The aim of *set the scene* is to 'start off on the right track' by introducing important activities for shaping the interaction from the beginning. Items in this domain express the educator's intention to help the learner improve; an acceptance that mistakes or omissions are expected while developing skills, arising from a growth mindset,[250] and involve the learner in a discussion about expectations and learning priorities for the session. However in our feedback videos, a comprehensive introduction was rarely seen.[57] In simulation-based education, a 'pre-brief' routinely occurs to explain goals, expectations and plans for the session and to foster a 'safe container'.[196] Work in the area of doctor-patient communication has highlighted the value of involving patients in developing the agenda, to set up a collaborative consultation.[227] In contrast, when someone does not know what is going to happen and feel they have little control over it, this promotes anxiety. Excessive anxiety interferes with attention, processing information and memory, all of which are important operations for learning.[98]

The next domain, *analyse performance,* focuses on the crucial step of assisting the learner to develop a clearer understanding of what the desired performance looks like and how their own performance compares with that.[3,14,18] Our qualitative analysis on evaluative judgement, published previously, contributed to revising these items in particular.[172] Items here highlight the value of clarifying key features of the target performance; grounding critique in specific examples to enhance understanding and credibility;[11,99,154] concentrating on 'did' not 'is' (otherwise, directing critique to personal

identity offers limited prospects for change and risks strong emotional reactions)[2,35] and prioritising discussion on a few points that are likely to be most useful for the learner, considering the learner's priorities and skill trajectory.[3,43] By endorsing aspects that the learner did correctly (or more correctly), the educator validates effective practice and confirms progress, which rewards effort, promotes intrinsic motivation and builds self-efficacy.[56,251] Additionally, clarifying the performance gap helps focus learners' attention on making improvements and paves the way for planning improvements.[1,54]

While *analyse performance* focuses on 'making sense', *plan improvement* deals with 'making use'.[26,249,252] Items in the *plan improvement* domain describe selecting important learning goals (such as addressing a significant error or responding to learner's request) and designing effective improvement strategies, tailored to the individual. Yet, studies report that action plans are often omitted.[57,117,119,171] Goal setting theory advocates that motivation, persistence and achievement are boosted when goals are clear and measurable (to determine progress), relevant and achievable (so valuable results compensate for effort) and with a deadline (to focus attention).[54,251]

The other two domains develop throughout a feedback conversation: *foster learner agency* and *foster psychological safety*. *Foster learner agency* incorporates themes of engagement, motivation and active learning.[54,113,152,251] According to social constructivism, as learners and educators propose, consider and hone ideas by building on each other's contributions, they co-create new insights and solutions.[20,43,253] The items describe ways to encourage learners to actively participate in an interactive learning conversation; to focus on developing their skills by reflecting on their performance, raising problems, asking questions and generating ideas for improvement.[113,152,157,193] When learners and educators critically analyse the learner's performance together, this offers a valuable opportunity for learners to refine their mental schemas about both the current task and broader learning skills, particularly evaluative judgement.[20,172,204] Strategies to support active learning permeate the other domains. For example, items in *analyse performance* encourage learner self-assessment and prioritising topics for discussion to avoid cognitive overload; and items in *plan*

improvements aim to ensure the learner understands the improvement strategy and rationale.

Foster psychological safety describes cultivating an environment in which *learner agency* can thrive. The importance of psychological safety stems from a learner's moment-to-moment dilemma, where engaging in productive learning behaviours entails the risk of an adverse outcome.[242] For example, if a learner asks a 'dumb question' or contests an educator's recommended strategy that the learner had tried to enact previously without success, this may expose undetected limitations in their knowledge and/or displease the educator.[31,194] Studies investigating learning and performance found that productive learning behaviours were common in ward teams with high psychological safety. These teams were characterised by three features: trust that co-workers had good intentions and were invested in each other's success; interest, acceptance and care for each other as individuals; and respect for each other's expertise.[192,193] These traits could be summed up as 'having someone's best interests at heart' and are embodied by collaboration. Based on principles identified in the literature and our own qualitative research study,[242] items depict ways educators can work with learners to nurture psychological safety; key themes include collaboration, respect, support and reducing the power gap.[29,196,197,242] An educator can promote the partnership by creating sustained opportunities for a learner to share their thoughts during learning activities (e.g. reflections, concerns or opinions) and respond in ways that demonstrate appreciation, curiosity, respect and support (e.g. showing compassion or suggesting ideas for overcoming challenges).[153,215,222] The inherent power imbalance between the defined roles of a supervisor/assessor and a learner may be moderated by educators demonstrating humility. In our feedback videos we saw educators acknowledge limitations in their own knowledge, assessment or advice; reveal difficulties they encountered during training;[31,228] endorse the need for life-long learning;[157] and appreciate the value of learners' contributions[215,222] Again, these themes are embedded in items across all the other domains.

Implications and future research

The FQI provides educators with a set of explicit behaviours designed to encourage a learner to collaborate in performance analysis and design of effective improvement strategies. Traditionally much advice for educators on feedback skills has contained principles such as 'work as allies', 'build trust' or 'be learner centred' but empirically-informed guidance on 'what might this look like' and how educators could help to cultivate these conditions, has been missing. We hope that by translating principles into actions and clearly articulating these standards, it will make it easier for educators to compare 'their work' (in this case, during feedback) with 'what is expected', just as learners do in trying to improve their clinical practice.[18] To support such professional development, we propose to create videos portraying feedback interactions to provide practical exemplars. These videos will involve actors performing fictional scenarios but informed by the authentic feedback videos. The FQI offers a framework that educators can use to analyse feedback encounters and trigger self-reflection. Clinicians could ask a colleague to observe their feedback practice, with learner consent, or instigate a 'video club' in which clinicians regularly discuss their own practice videos.[254] In either case, the critique could be stimulated by items on the FQI, rather than 'gut feels' about whether or not a feedback session was effective.[255] This could assist everyday clinical educators to gain expertise and confidence in promoting effective feedback interactions.

We plan to undertake further testing of the FQI, including feasibility and 'think aloud' testing,[256] and psychometric analysis using a larger sample, which may lead to further refinement. The FQI offers future opportunities to systematically analyse feedback to identify which educator behaviours, or combinations, have the greatest influence on learner outcomes. After all, the ultimate test for feedback quality is its effect.[10] This could identify a smaller number of the most useful behaviours, to create a 'mini-FQI' that is easy for everyday clinicians to adopt. Additionally, Rasch analysis of a finalised FQI could provide insights on a developmental trajectory in feedback proficiency, as Rasch analysis orders items (and therefore behaviours) from easiest to hardest. This could provide support for sequencing of educator training (analogous to a child learning to count, add, then multiply during mathematical skills progression).

Strengths and limitations of research

The strengths of this research lie in the rigorous development of the Feedback Quality Instrument. Phase 1, previously published, involved extensive literature searching for empirical evidence and Delphi processes with an expert panel to achieve consensus on a provisional feedback instrument.[52] Phase 2, detailed here, involved administering the provisional instrument to analyse authentic feedback episodes with diverse health professionals, then refining it based on usability testing, psychometric analysis and parallel qualitative research on psychological safety and evaluative judgement.[242]

There are a number of limitations to our research. Clinicians and students who volunteered to participate may not have been representative of supervising clinicians in general. Videoing feedback interactions may have influenced participant behaviour. Inconsistencies in observed ratings may be improved by item refinements and rater training using exemplars, calibration training and an instrument manual. The data set size was at the lower acceptable limit and a larger data set would enhance confidence in results from both Rasch and exploratory factor analysis. The FQI was designed in one country, involving multiple academics and clinicians across three states, and tested within one major healthcare network. Therefore, how applicable the instrument is to different countries and contexts is unknown.

7.8 Conclusions

This study resulted in the Feedback Quality Instrument, ready-for-use in clinical practice. The FQI contains five domains portraying core concepts that constitute high quality feedback. Three domains occur sequentially, *set the scene*, *analyse performance* and *plan improvement* and two flow throughout a feedback encounter, *foster psychological safety* and *foster learner agency*. This instrument offers educators a set of explicit descriptions of useful behaviours to guide workplace feedback. By orientating educators to what learner-centred feedback 'could look like', we hope it promotes conversations that help learners to develop.

Chapter 7 Supplementary information

Methods

S7.1 Administering the provisional instrument to evaluate educators' feedback practice

The raters were all female health professionals with senior education and research roles (two doctors and four physiotherapists), including three with clinical roles in different units within the health service (two doctors and one physiotherapist).

As a pilot within this study, raters independently analysed three videos and then discussed utilising the provisional instrument and any related difficulties. Strategies to address the two problems identified were developed and implemented as follows. *Item 2: The educator offered to discuss the performance as soon as practicable* would occur prior, not during a feedback conversation, so it was removed from the instrument. Hence 24 items were scored and the total rating score range was 0 to 48. For *Item 10: The educator acknowledged and responded appropriately to emotions expressed by the learner,* it was decided to rate this as '2' (done consistently) if either a) the learner displayed signs of emotion, including verbal indicators (e.g. defensive language); paraverbal indicators (e.g. slow monotone speech suggesting feeling disheartened); or nonverbal indicators (e.g. facial flushing suggesting embarrassment) and the educator responded appropriately, or b) if the learner's demeanour throughout suggested a comfortable emotional equilibrium.

S7.2 Usability analysis of the provisional instrument

Raters commented on ease of instrument use; item overlaps (when more than one item addressed similar educator behaviours) or gaps (when pertinent educator behaviours were not captured by items); item interpretation including phrasing that was too complex, could be interpreted in various ways, or did not effectively capture relevant observable behaviours (for example, if descriptions were either too nebulous or restrictive); and rating category interpretation and application.[241]

S7.3 Multifaceted Rasch Model analysis (MFRMA)

'Feedback proficiency' cannot be directly measured as it is a psychosocial construct characterised by certain behaviours, whereas height or weight can be directly measured using a ruler or scales respectively. The provisional feedback instrument contained items that described observable behaviours, rated using a Likert scale. However the 'distance' between one rating category and the next on a Likert scale may not be uniform, and the items may vary in difficulty. Hence a total score, resulting from the sum of individual item scores, may not accurately reflect differences in overall proficiency levels between educators. Rasch models, based on item response theory, describe the characteristics of assessment instruments used to convert these type of ordinal data (i.e. Likert scale) into a linear interval performance measure.[257-259] In this case, the model computes the probability that an educator will demonstrate a specific behaviour recommended in high quality feedback discussions, assuming it is solely determined by the educator's proficiency and the difficulty of the behaviour described in the item (mastery level). When an educator correctly demonstrates a specific behaviour, there is a high probability that they will correctly demonstrate behaviours described in all easier items. Items can be ranked from easiest to most difficult and this order remains consistent for educators with different proficiency levels. This allows 'proficiency' and 'item difficulty' to be represented on the same linear interval scale, which creates a measurement scale (or 'ruler') that estimates feedback proficiency. Several benefits arise from this. A linear interval scale allows differences between proficiency scores to be more accurately quantified (for example, when an educator improves their proficiency level over time or to compare the proficiency levels of different educators) and typically, the data are more suitable for parametric analysis than ordinal data.

In this study, a Multifaceted Rasch Model was used, which took account of the influence of rater severity, item difficulty, rating categories and educator proficiency on the score (each called a 'facet').[232,260] As the main purpose was to refine the provisional instrument, the analysis focused on investigating how closely the observed item ratings matched those expected by the model. This was primarily assessed using model fit statistics; items with sufficient 'misfit' to the model may degrade the instrument's

function as a measurement scale. However, misfit only highlights items for review; it does not necessitate removal or revision. Item difficulty, educator proficiency and rater leniency were presented on the same linear interval scale (the 'feedback quality ruler') with 'logits' as the unit of measurement, illustrated in a variable map. This allowed comparisons between facets and components of facets (for example, individual items). In particular, the variable map provided information on whether the provisional instrument had sufficient coverage of items across the full range of feedback proficiency. The instrument's power to discriminate between different proficiency levels demonstrated by educators was investigated using a separation reliability index. The MFRM analysis was conducted using ConQuest Generalised Item Response Modelling Software, Version 4, 2015 (Australian Council for Educational Research, Camberwell, Victoria).

Model fit statistics

Fit statistics indicate how closely observed data matched expected data generated by the MFRM model. Residuals are the difference between the observed and expected item ratings. The fit statistics included i) mean square of the residuals (MNSQ) with a 95% confidence interval and ii) a normalised equivalent ('T' statistic).[234,261] The MNSQ indicates the size of any misfit and the T statistic indicates the likelihood of misfit. Fit statistics are interpreted in a similar manner for all facets (items, raters and rating categories). A residual represents 'effect on the instrument score, apart from the trait' and denotes 'noise', as opposed to the 'signal' indicating the trait. Residuals can include an assortment of influences such as item phrasing; personal factors influencing participants' performance or raters' judgments, such as personal experience or fatigue; or different contexts. In other words, 'noise' describes any contribution that detracts from the desired consistent interpretation of the item content and rating criteria across different contexts.

The mean residual for all parameters is set to 1.0 and has a range from zero to infinity. The MNSQ also indicates the proportion of variance in observed responses from that expected by the model i.e. MNSQ of 1.30 indicates 30 percent more variance in

responses than expected by the model and 0.77 is 30 percent less variance than expected by the model. In general, MNSQ values between 0.5 to 1.5 indicate acceptable fit. For clinical observations, MNSQ between 0.5 to 1.7 are reasonable.[261,262] High MNSQ values (known as underfit) indicate observed data are less predictable (i.e. more variable) than expected. MNSQ values between 1.5 and 2 are unproductive for measurement construction but not distorting; MNSQ values greater than 2 distort the instrument's measurement system.[262,263] T values greater than 2 indicate observed data are unlikely to fit the Rasch model (i.e. misfit is likely). On the other hand, low MNSQ values (known as overfit) below 0.5 indicate observed data are too predictable. T values greater than -2 indicate observed data are unlikely to fit the Rasch model. Overfit does not usefully contribute to the instrument's measurement system but it does not distort it. The main concern is that overfit can artificially inflate the instrument's separation reliability index.[259]

'Unweighted' fit (sometimes called 'outfit') is sensitive to unexpected observed data involving participants with remote proficiency levels, for example, far away from the difficulty level of an item. Typical reasons for this include a careless mistake (e.g. when a participant with high overall proficiency does not demonstrate a behaviour described in an easy item) or subspecialty expertise (e.g. when a participant has a much higher proficiency in one particular area so they demonstrate a behaviour described in a difficult item, well above their overall proficiency level). 'Weighted' fit (sometimes called 'infit') corrects for this by giving more weight to item ratings for participants with proficiency levels close to the difficulty of that item. For example, when there is a large weighted fit value, this means the component (for example, an item) is not functioning well at the location where it is most discriminating, which is more serious. Typical reasons for this include the item targets an inconsequential behaviour, ambiguous item phrasing, or sparse data (more sensitive to unexpected observations). On the other hand, when there is a large value for unweighted fit but not weighted fit, this is less serious as it is most likely due to idiosyncratic events.

In brief, misfit was used to identify specific facet components (items, raters or rating categories) for which observed data was more erratic than expected by the MFRM. Misfit was indicated by weighted MNSQ greater than 1.5 (suggesting items may not be usefully contributing the measurement system), particularly greater than 2 (suggesting items may be distorting the measurement system) and T greater than 2 (suggesting misfit was likely).

As the primary aim of this analysis was to refine the provisional instrument, a sensitivity analysis was conducted to better isolate potential problems due to the items themselves, by controlling for rater and rating scale problems that may have a substantial influence on item fit. When misfit was identified, this led to a thorough review of potential reasons, which informed consideration of subsequent modifications.[262,263]

Raters and rating scale fit statistics

Rater and rating category fit statistics provided information about how raters interpreted and applied items and rating categories. Rater underfit (MNSQ greater than 2 and T greater than 2) indicated raters who rated educators' practice more erratically than expected by the Rasch model. Overfit (MNSQ less than 0.5 and T below -2) indicated raters with little variability in their ratings for educators across the range of feedback proficiency. Rating category fit statistics were used to identify inconsistent selection of a rating category at the expected proficiency level. Typical reasons for rating category misfit includes rating category problems, such as too many rating categories or vague descriptions of rating categories so it is hard to differentiate between them; or item phrasing problems that result in inconsistent selection or inapplicable rating categories.

Separation reliability statistics

Separation reliability statistics for each facet provide an indication of discriminatory power. It is reported with a chi-square statistic (X^2), to compare the distribution of components within each facet, the degrees of freedom (df) and a probability level (P).

Separation reliability statistics are interpreted in a similar way as Cronbach's alpha, with a range from 0 to 1, and larger values indicating greater consistency. The person separation reliability indicates the instrument's ability to discriminate educators with different levels of feedback proficiency: 0.8 is acceptable and indicates ability to discriminate at least 2 different levels of feedback proficiency and 0.9 indicates ability to discriminate at least 4 different levels.[264] Multiple raters, as in this study, may result in higher reliability compared to single raters, as typically occurs in routine instrument administration.

Variable map including item distribution

A variable map was used to show the spread of components within each facet on the same linear scale, including items (easy to difficult), participants (low to high proficiency) and raters (lenient to severe). In particular, the distribution of items was inspected. Each item can be considered as a 'mark' on the 'feedback ruler', so an ideal instrument would have items distributed across the full range of feedback quality, from very easy to very difficult items. If there were no items to measure the lowest or highest levels of proficiency, this could lead to 'floor' or 'ceiling' effects, respectively. There should be no substantial span without items; the instrument would be less able to precisely differentiate between proficiency levels, as it would not be generating as much information in that region. Conversely multiple items in a small span may lead to redundancy and unnecessary level of precision. Hence examination of the variable map enabled identification of gaps, redundancy and propensity to produce floor or ceiling effects.

S7.4 Exploratory factor analysis

Evaluating suitability of data for EFA

We assessed the suitability of the provisional instrument ratings data for EFA using Bartlett's test of sphericity[265] ($p < 0.05$ indicating suitability) and Kaiser-Meyer-Olkin measure of sampling adequacy[266] (a value greater than 0.6 is sufficient,[248] with higher levels indicating increasingly compact correlation patterns with 0.71-0.8 described as 'good', 0.81-0.9 as 'very good' and over 0.9 as 'excellent').[267]

Factor extraction

After determining that the data were suitable, principal components analysis (one EFA approach) was used to determine the most appropriate number of factors.[243] This involved balancing competing aims to summarise the data using as few factors as possible but also to account for as much variability in the data, explaining the inter-relationships between the items. The first factor extracted explains the most variance within the data set and each subsequent factor extracts explains less variance. The objective is to account for as much variance in the data with as few factors as possible. We determined the number of factors to analyse further by considering factor eigenvalues, the total variance extracted, the variance extracted by each factor, the scree plot and parallel analysis. An eigenvalue indicates the amount of total variance across all items explained by the factor, and typically only factors with eigenvalues greater than 1.0 are selected. Variance refers to the variability in observed results for items within a cluster that is explained by the underlying factor (that is, the underlying concept shared by the items). It is recommended that retained factors extract over 5% variance, with cumulative variance between 50 to 75%.[244] The scree plot is a graph that plots the eigenvalue against each factor in increasing numerical order. [268] It is used to identify factors that account for the greatest variance, which occur on the steep section of the curve before it flattens out. Parallel analysis was used to identify factors with eigenvalues greater than the corresponding mean eigenvalues calculated from 100 random data sets of the same size, with the aim of excluding factors with eigenvalues no greater than those achieved in a random data set.[269]

After selecting the number of factors to extract, factor rotation was conducted to assist with interpretation of item clusters. Factor rotation can offer alternative factor solutions, by rotating items in multidimensional space while maintaining inter-relationships. The aim is to produce a 'simple structure' in which each item strongly loads only onto one factor (typically, at least 0.4).[243] Factor loading signifies the correlation between an item and a factor, which reflects how strongly the underlying concept (represented by the factor) influences the observed result for that item. Selecting the best arrangement of factors and item clusters involved making decisions

based on both the statistical analyses (ideally each item strongly loads onto one factor only) and theoretical considerations (maintaining the theoretical foundations of the instrument).[244] A direct oblimin technique was used; this 'oblique approach' makes no assumptions about relationships between factors. In addition, a varimax technique was used when factors were not closely related (when factor correlations were less than 0.3); this 'orthogonal approach' assumes the underlying constructs are independent. The varimax rotation aims to make the factors as distinct as possible, in which each factor only has items that strongly load onto it and no other factor, by amplifying the shared variance between the items and factors. Communality values provide information on the variance of each item explained by the factor, with values less than 0.3 indicating an item does not fit in well with the other items in the factor.[243]

Once the best arrangement was selected, the factors were named according to the shared characteristic, indicated by the strongest loading items. The EFA was conducted using IBM SPSS Statistics for Windows, V25.0, 2017 (IBM Corp, Armonk, NY). (For more details, see supplementary information: Sections 3.1 and 3.2).

Results

S7.5 Administering the provisional instrument

Feedback videos and health professional participants

The feedback session was typically recorded by setting up a computer or smart phone to record it. The discussion was based on the learner's performance of a specific clinical task in 25 (69.4%) videos and across a clinical attachment in a mid/end-of-attachment appraisal in 11 (30.6%) videos. Most assessments were formative but some were summative contributions to longitudinal training evaluations. Participants included 34 educators and 35 learners (with two educators and one learner involved in two videos each). Educator characteristics included 18 (52.9%) women and 16 (47.1%) men; with 26 medical (six physicians, four surgeons, three anaesthetists, three emergency physicians, three psychiatrists, three obstetrician-gynaecologists, two radiologists, one paediatrician and one opthalmologist), 4 nursing and 4 physiotherapy health

professionals. Experience levels included 14 (41.2%) educators with five years or less experience supervising learners, 11 (32.3%) with six to ten years, and 9 (26.5%) with more than ten years. The learner characteristics included 23 (65.7%) women and 12 (34.3%) men; in the same disciplines as their educators; and the following experience levels: 9 (25.7%) students, 9 (25.7%) clinicians with five years or less post-qualification, 15 (42.9%) clinicians with 6 years or more post-qualification and 2 (5.7%) senior clinicians.

Item ratings frequency data

The frequency data for item ratings are shown in Table S7.1.

Table S7.1: Frequency data for item ratings, resulting from administration of the provisional instrument to analyse feedback videos

Item	Rating = 0	Rating = 1	Rating = 2
1	25	23	125
3	158	14	1
4	137	29	7
5	148	22	3
6	32	40	100
7	135	18	19
8	57	44	70
9	65	48	58
10	11	18	142
11	8	43	122
12	133	21	19
13	62	54	57
14	31	34	108
15	30	46	97
16	21	56	95
17	9	31	66
18	11	28	68
19	53	30	24
20	8	28	135
21	57	55	60
22	67	59	46
23	139	21	13
24	152	13	7
25	165	6	1

Rating scale: 0 = not seen; 1 = done somewhat or sometimes; 2 = done consistently

Usability analysis of the provisional instrument

Raters reported that it was demanding to rate 24 separate items in the provisional instrument (See Figure 7.2 in the main article for items). For specific items, they noted the following issues: Item 1 could be interpreted as 'done' in the context of a workplace based assessment; Items 1 and 17 targeted similar educator behaviours; Items 7, 12 and 13 included the phrase 'the educator asked' which either did or did not occur, which resulted in category rating 1 (done somewhat or sometimes) becoming inapplicable;

Item 11 was too nebulous; and Items 17-19 were not generally applicable during a feedback discussion, as they had been designed as a subset of items to be used when an educator raised performance concerns but this did not occur in all interactions. Hence items 1, 7, 11, 12, 13, 17-19 were flagged for review.

S7.6 Multifaceted Rasch Model analysis

Item analysis including sensitivity analysis

Several items demonstrated misfit (weighted MNSQ > 1.5 and T > 2): items 5*, 6*, 8, 14, 15, 16* and 23 (* denotes MNSQ > 2, indicating more serious misfit). The detailed item fit data are presented in Table S7.2.

In the sensitivity analysis, designed to isolate problems due to items themselves, by reducing the effect of rater or rating category misfit, items 5*, 6, 14 and 23 demonstrated misfit (weighted MNSQ > 1.5 and T > 2; * denotes MNSQ > 2, indicating more serious misfit). The sensitivity analysis involved reducing the influence of rater and rating category problems on item fit: data were reanalysed after removing data from raters 2 and 6 (as rater 2 showed misfit and rater 6 was markedly severe) (See additional details below under 'rater analysis' and Table S3) and combining rating categories 1 and 2 to create one category 'done, at least partially' (as rating category 1 showed misfit) (See additional details below under 'rater category analysis' and Table S4). The provisional instrument had three rating categories: 0= 'not seen', 1 = 'done somewhat or sometimes', 2 = 'done consistently'.

Therefore, all these items were flagged for review, particularly those items which demonstrated misfit in the sensitivity analysis and especially Item 5 which demonstrated more serious misfit (with MNSQ > 2).

Overfit

The initial MFRM analysis demonstrated overfit for Item 11 (MNSQ < 0.5 and T > - 2) in weighted and unweighted fit but none in the sensitivity analysis.

Table S7.2: Item fit from multifaceted Rasch model analysis of the provisional instrument.

Item	Initial analysis					Sensitivity analysis for item fit				
		Unweighted fit		Weighted fit			Unweighted fit		Weighted fit	
	Estimate in logits (standard error)	MNSQ (95% confidence intervals)	T	MNSQ (95% confidence intervals)	T	Estimate in logits (error)	MNSQ (95% confidence intervals)	T	MNSQ (95% confidence intervals)	T
1	-1.62 (0.14)	1.25 (0.54, 1.46)	1.0	1.16 (0.49, 1.51)	0.6	-2.38 (0.19)	1.62 (0.54, 1.46)	2.3	1.41 (0.36, 1.64)	1.2
3	2.62 (0.24)	1.21 (0.54, 1.46)	0.9	1.22 (0.27, 1.73)	0.7	2.91 (0.24)	1.59 (0.54, 1.46)	2.2	1.58 (0.19, 1.81)	1.3
4	1.63 (0.16)	0.91 (0.54, 1.46)	-0.3	0.91 (0.44, 1.56)	-0.2	2.02 (0.18)	1.51 (0.54, 1.46)	2.0	1.49 (0.39, 1.61)	1.5
5	2.07 (0.19)	2.60 (0.54, 1.46)	4.8	2.24 (0.38, 1.62)	3.0	1.98 (0.18)	2.41 (0.54, 1.46)	4.4	2.01 (0.40, 1.60)	2.7
6	-1.13 (0.12)	1.93 (0.54, 1.46)	3.2	2.31 (0.52, 1.48)	4.0	-1.07 (0.14)	1.44 (0.54, 1.46)	1.7	1.74 (0.52, 1.48)	2.6
7	1.34 (0.14)	1.97 (0.54, 1.46)	3.3	1.38 (0.47, 1.53)	1.3	2.38 (0.21)	2.80 (0.54, 1.46)	5.3	1.71 (0.32, 1.68)	1.8
8	-0.43 (0.11)	1.63 (0.54, 1.46)	2.3	1.71 (0.54, 1.46)	2.6	- 0.39 (0.13)	1.39 (0.54, 1.46)	1.5	1.48 (0.55, 1.45)	1.9
9	-0.18 (0.11)	1.47 (0.54, 1.46)	1.8	1.39 (0.54, 1.46)	1.6	0.11 (0.13)	1.54 (0.54, 1.46)	2.1	1.47 (0.55, 1.45)	1.9
10	-2.33 (0.17)	1.03 (0.54, 1.46)	0.2	1.37 (0.43, 1.57)	1.2	-2.59 (0.21)	0.73 (0.54, 1.46)	-1.2	1.29 (0.31, 1.69)	0.9
11	-1.89 (0.15)	0.41 (0.54, 1.46)	-3.2	0.41 (0.47, 1.53)	-2.8	-2.49 (0.20)	0.80 (0.54, 1.46)	-0.8	0.85 (0.34, 1.66)	-0.4
12	1.28 (0.14)	1.14 (0.54, 1.46)	0.6	0.93 (0.48, 1.52)	-0.2	1.54 (0.16)	1.29 (0.54, 1.46)	1.2	1.05 (0.46, 1.54)	0.2
13	-0.21 (0.11)	0.77 (0.54, 1.46)	-1.0	0.75 (0.54, 1.46)	-1.1	-0.30 (0.12)	1.16 (0.54, 1.46)	0.7	1.12 (0.55, 1.45)	0.6
14	-1.26 (0.12)	1.72 (0.54, 1.46)	2.6	1.66 (0.51, 1.49)	2.3	-1.54 (0.15)	1.79 (0.54, 1.46)	2.8	1.66 (0.48, 1.52)	2.2
15	-1.11 (0.12)	1.34 (0.54, 1.46)	1.4	1.60 (0.52, 1.48)	2.1	-1.49 (0.15)	0.97 (0.54, 1.46)	-0.0	1.23 (0.49, 1.51)	0.9
16	-1.22 (0.12)	1.67 (0.54, 1.46)	2.4	1.95 (0.51, 1.49)	3.1	-1.26 (0.14)	1.25 (0.54, 1.46)	1.1	1.35 (0.51, 1.49)	1.3
17	-1.24 (0.16)	1.11 (0.52, 1.48)	0.5	1.21 (0.44, 1.56)	0.8	-1.19 (0.18)	0.92 (0.52, 1.48)	-0.3	0.96 (0.43, 1.57)	-0.0
18	-1.26 (0.16)	1.04 (0.52, 1.48)	0.3	1.16 (0.44, 1.56)	0.6	-1.47 (0.19)	1.13 (0.52, 1.48)	0.6	1.19 (0.39, 1.61)	0.7
19	0.52 (0.14)	1.24 (0.52, 1.48)	1.0	1.33 (0.50, 1.50)	1.2	0.63 (0.16)	1.06 (0.52, 1.48)	0.3	1.14 (0.51, 1.49)	0.6
20	-2.19 (0.16)	1.77 (0.54, 1.46)	2.7	1.40 (0.44, 1.56)	1.3	-2.60 (0.21)	1.77 (0.54, 1.46)	2.7	1.40 (0.31, 1.69)	1.1
21	-0.30 (0.11)	1.00 (0.54, 1.46)	0.1	0.99 (0.54, 1.46)	0.0	-0.20 (0.12)	1.17 (0.54, 1.46)	0.8	1.13 (0.55, 1.45)	0.6
22	-0.01 (0.11)	1.11 (0.54, 1.46)	0.5	1.07 (0.54, 1.46)	0.4	-0.20 (0.12)	1.04 (0.54, 1.46)	0.3	1.05 (0.55, 1.45)	0.3
23	1.53 (0.15)	1.27 (0.54, 1.46)	1.1	1.82 (0.45, 1.55)	2.4	1.93 (0.18)	1.06 (0.54, 1.46)	0.3	1.72 (0.41, 1.59)	2.1
24	2.10 (0.19)	1.04 (0.54, 1.46)	0.2	1.50 (0.37, 1.63)	1.4	2.39 (0.21)	0.80 (0.54, 1.46)	-0.8	1.27 (0.32, 1.68)	0.8
25	3.29 (0.34)	1.00 (0.54, 1.46)	0.1	0.89 (0.09, 1.91)	-0.1	3.27 (0.83)	0.56 (0.54, 1.46)	-2.2	0.82 (0.08, 1.92)	-0.3

Shading: ▢ MNSQ = 1.5 to 2 and T = greater than 2; ▢ MNSQ = greater than 2 and T = greater than 2.
Abbreviations: MNSQ = mean square of the residuals; T = T statistic (normalised equivalent).

Item separation reliability

The item separation reliability was 0.99, X^2 =1958, df 23, P <0.000. This indicated the item difficulty estimates were significantly different from each other and covered a range of levels on the underlying scale.

Rater and rating category analysis

Rater fit statistics revealed rater 2 showed severe misfit (weighted MNSQ = 4.7; T = 8.6) and rater 5 showed misfit (weighted MNSQ = 2.3; T = 2.3) (misfit indicated by weighted MNSQ > 2 and T > 2). As rater 5 analysed only 10/36 (28%) of the videos, the sparsity of data limits certainty in this result. Rater 6 was markedly more severe (severity estimate was 1.45 logits) than the remaining rater group, which had a narrow range in severity estimate from -0.81 to 0.13 logits. Rater separation reliability was 0.957, X^2 = 228.20, df 5, P < 0.000, which confirmed the raters did not rate to a common standard with Rater 6 demonstrating the largest difference in comparison with the average (see Variable map, Figure 6). Rating category fit statistics revealed rating category 1 showed misfit with MNSQ = 2.1 and T = 3.5 (weighted MNSQ > 2 and T > 2). Therefore, potential reasons for rating category 1 misfit were considered, including i) this rating category captures the partial demonstration of a recommended behaviour and the description 'somewhat', hence this not as well defined as 'done' or 'not done', which may lead to less consistent use, or ii) this rating category may not be applicable for all the items, for example any item that describe a behaviour that can only be done / not done, such as 'asked'.

The detailed rater fit and rater severity data are shown in Table S7.3. The detailed rating category data are shown in Table S7.4.

Table S7.3: Rater severity and rater fit from multifaceted Rasch model analysis of the provisional instrument, involving 6 raters ranked in order of severity (most severe rater at the top).

		Unweighted fit		Weighted fit	
Rater	**Severity estimate in logits (standard error)**	**MNSQ (95% confidence intervals)**	**T**	**MNSQ (95% confidence intervals)**	**T**
6	1.45 (0.09)	2.50 (0.40, 1.60)	3.6	1.88 (0.38, 1.62)	2.3
5	0.13 (0.11)	2.32 (0.12, 1.88)	2.3	2.29 (0.13, 1.87)	2.3
3	-0.18 (0.06)	1.72 (0.54, 1.46)	2.6	1.67 (0.54, 1.46)	2.4
4	-0.19 (0.06)	1.49 (0.54, 1.46)	1.9	1.48 (0.54, 1.46)	1.8
2	-0.40 (0.06)	4.68 (0.54, 1.46)	8.6	4.67 (0.54, 1.46)	8.6
1	-0.81 (0.06)	1.81 (0.53, 1.47)	2.8	1.77 (0.53, 1.47)	2.7

Abbreviations: MNSQ = mean square of the residuals; T = T statistic (normalised equivalent).

Table S7.4. Rating category results from multifaceted Rasch model analysis of the provisional instrument.

		Unweighted fit		Weighted fit	
Rating category	**Category estimate in logits (error)**	**MNSQ (95% confidence intervals)**	**T**	**MNSQ (95% confidence intervals)**	**T**
0		1.34 (0.54, 1.46)	1.4	1.30 (0.53, 1.47)	1.1
1	-0.074 (0.049)	2.19 (0.54, 1.46)	3.9	2.06 (0.53, 1.47)	3.5
2	0.074	1.23 (0.54, 1.46)	1.0	1.21 (0.53, 1.47)	0.8

Abbreviations: MNSQ = mean square of the residuals; T = T statistic (normalised equivalent).

Variable map

The variable map is presented in Figure 7.6 (main article). The Rasch 'ruler' with the scale in logits (unit of measurement) is shown on the far left, representing the range of 'feedback proficiency'. For each facet, the lowest levels are shown at the bottom of the variable map (indicated by the lowest logit estimates), with increasing levels towards the top of the variable map. Educators are shown on the left side of the left vertical hashed line, positioned according to their proficiency level (each 'X' represents 0.3 participant; this provides a slight distribution incorporating each participant's estimate and standard error). Items are shown on the right side of the right vertical hashed line, located according to the average relative difficulty of the item (within each item, score categories (0,1,2) are distributed over a wider range, which is not shown here).

S7.7 Exploratory factor analysis

Principal components analysis

The data were found to be suitable for factor analysis (Bartlett's test of sphericity was statistically significant with p = 0.000 and the Kaiser-Meyer-Olkin level was 0.825, exceeding the recommended value of 0.60).

Principal components analysis of the provisional instrument indicated that between three to five factors should be retained for further analysis. This was based on the following: seven factors had eigenvalues over one, four to seven factors accounted for 54.8 to 69.0% total variance,[244] parallel analysis revealed three factors had eigenvalues greater than corresponding criterion values for a randomly generated data set of the same size[269] and the scree plot displayed a sharp break point after the second factor.[268] See Table S7.5 for the principal components analysis, Figure S1 for the scree plot and Table S7.6 for the parallel analysis.

Rotation using a direct oblimin approach was conducted on three-factor, four-factor and five-factor extractions. The four-factor solution with oblimin rotation showed a simple structure: each factor had multiple items, each with high loadings on to it, which were theoretically aligned together, and no substantial loadings onto other factors; there

were only a few items with only a low loading on one item or crossloadings (that is, with similar substantial loadings on multiple factors) (See Table S7.7). Communality values were low for items 5 and 25, indicating these items did not strongly share the same characteristics as the other items in their clusters). Rotation with a varimax approach gave a similar result. The four factors were named according to the highest loading items: factor one: *analyse performance*, factor two: *foster learner agency*, factor three: *plan improvement* and factor four: *set the scene*.

The four-factor solution was not perfect. Items 10 and 11 were both split between two factors and were not theoretically congruous with either of them. However, in a five-factor solution with Oblimin rotation, Items 10 and 11 strongly loaded together to form a fifth factor (See Figure 7.2 in main article for item list). Therefore, it was decided to create a fifth factor based on these items. Additionally in the four-factor solution, Item 5 did not load onto any factor, Item 7 loaded onto two factors, and Item 25 only had a low loading onto a factor.

Table S7.5: Results of the principal components analysis of the provisional instrument.

Total Variance Explained							
Component	Initial Eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings
	Total	Percentage of variance	Cumulative percentage	Total	Percentage of variance	Cumulative percentage	Total
1	7.60	31.7	31.7	7.60	31.7	31.7	5.5
2	2.58	10.7	42.4	2.58	10.7	42.4	4.8
3	1.54	6.4	48.8	1.54	6.4	48.8	3.8
4	1.44	6.0	54.8	1.44	6.0	54.8	1.8
5	1.27	5.3	60.1	1.27	5.3	60.1	2.3
6	1.13	4.7	64.8	1.13	4.7	64.8	1.8
7	1.01	4.2	69.0	1.01	4.2	69.0	1.3
8	.88	3.7	72.7				
9	.80	3.3	76.0				
10	.74	3.1	79.1				
11	.65	2.7	81.8				
12	.60	2.5	84.3				
13	.53	2.2	86.5				
14	.46	1.9	88.4				
15	.41	1.7	90.1				
16	.40	1.7	91.8				
17	.37	1.5	93.3				
18	.33	1.4	94.7				
19	.30	1.3	96.0				
20	.27	1.1	97.1				
21	.20	.9	97.9				
22	.18	.8	98.7				
23	.17	.7	99.4				
24	.15	.6	100.0				

Figure S7.1: Scree plot showing eigenvalues of consecutive extracted factors, from principal components analysis of the provisional instrument.

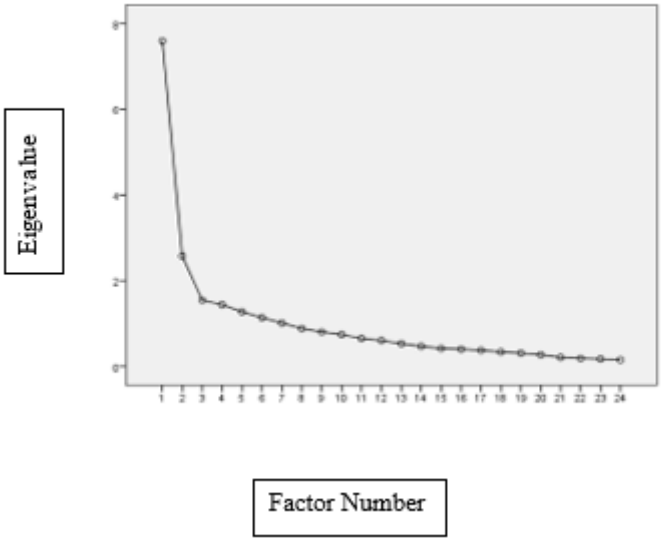


Table S7.7: Pattern and structure matrix for principal components analysis with oblimin rotation of a four-factor solution.

Item	Pattern coefficients				Structure coefficients				Communalities
	Factor				Factor				
	1	2	3	4	1	2	3	4	
Item 1	**.807**	.019	-.015	.005	**.809**	.234	.195	.108	.655
Item 20	**.798**	.051	-.101	-.075	**.777**	.232	.101	.019	.621
Item 14	**.777**	-.170	.100	.190	**.781**	.091	.292	.275	.672
Item 15	**.768**	-.113	.193	.044	**.792**	.139	.373	.152	.671
Item 17	**.705**	.115	.041	.153	**.766**	.339	.267	.269	.631
Item 18	**.665**	.225	-.017	.156	**.741**	**.427**	.222	.276	.627
Item 11	**.627**	**.413**	-.173	-.107	**.680**	**.528**	.053	.017	.633
Item 16	**.614**	-.096	.254	.101	**.665**	.137	**.406**	.202	.519
Item 6	.051	**.797**	.032	-.021	.270	**.813**	.202	.127	.665
Item 8	.081	**.770**	.188	-.007	.335	**.829**	.362	.164	.732
Item 9	.007	**.709**	-.002	.235	.226	**.751**	.179	.357	.617
Item 19	.077	**.653**	.321	-.111	.319	**.719**	**.454**	.060	.633
Item 13	.358	**.522**	.103	.038	**.530**	**.646**	.305	.189	.566
Item 12	-.116	**.519**	.264	.115	.105	**.561**	.357	.230	.400
Item 10	.366	**.467**	-.148	-.210	**.427**	**.499**	.006	-.106	.411
Item 23	-.074	.233	**.727**	.146	.192	.385	**.778**	.289	.683
Item 24	-.116	.234	**.602**	.057	.107	.334	**.629**	.175	.455
Item 22	.264	.128	**.553**	-.108	**.426**	.292	**.630**	.033	.495
Item 21	.186	.132	**.523**	.238	.384	.328	**.634**	.365	.533
Item 25	.067	-.073	**.474**	-.158	.148	.014	**.452**	-.089	.237
Item 4	.102	.037	-.065	**.734**	.189	.177	.081	**.743**	.566
Item 3	.160	-.062	-.006	**.640**	.223	.089	.121	**.649**	.445
Item 7	-.173	**.465**	-.161	**.566**	-.017	**.483**	-.024	**.599**	.571
Item 5	.098	-.047	.178	.229	.160	.055	.229	.261	.114

Footnote: Major loadings for each item in bold

S7.8 Refinements to the provisional instrument based on qualitative analyses

The two qualitative analyses conducted on the feedback video data provided practical insights into ways that educators in our study worked with learners to foster their psychological safety and evaluative judgement. Below are brief summaries of the main findings of these studies that informed refinement of the provisional instrument.

Psychological safety: Thematic analysis of our feedback video transcripts explored ways psychological safety manifested during feedback sessions in clinical practice, published previously.[242] We inferred psychological safety when learners became increasingly candid and engaged in productive learning behaviours during the discussion, then we looked for corresponding educator behaviours. We created themes that described how educators had collaborated with learners to foster psychological safety. Based on our inductive analysis, we found educators seemed to foster psychologically safety by 1) setting expectations for learners to participate from the start; 2) collaborating with learners by demonstrating respect, compassion and assistance, and actively reducing the power gap; 3) focusing on continuing improvement, consistent with Dweck's 'growth mindset' in which limitations and mistakes are expected during skill development and are seen as learning opportunities[250] and 4) valuing learners' contributions to honest dialogue, by exploring learners' perspectives, listening attentively and thoughtfully responding. These findings contributed to new items as well as revising items in the provisional instrument.

Evaluative Judgement: Thematic analysis of our feedback video transcripts explored what opportunities educators provided for learners to develop their evaluative judgement, published previously.[172] We reported three themes 1) *Educator's invitations: I want to know what you think about the quality of your work*, which described how educators encouraged learners to share genuine evaluations of their work, often using follow-up probes for the learner to offer deeper reflections; 2) *Good work looks like this.* Educators commonly clarified standards, often using examples from their own practice; 3) *Calibration of the learner's judgement through dialogue*. This

theme highlighted how important it was educators to respond to learners' comments about their work, to confirm or correct them, to fine tune their judgements.

Details of the item refinements, including new items, are outlined in detail in Supplementary information Table S7.8.

Supplementary information S7.8

The FQI evaluates the quality of an educator's contribution to face-to-face verbal feedback, following observation of a learner performing a task in clinical practice.

The 'concept catcher' encapsulates the essence of each item, to assist the user; this appears in grey italics.

Desirable criteria

The instrument overall should:

- be comprehensive (minimise gaps)

- be parsimonious (minimise overlap)

Each item should:

- describe an observable educator behaviour

- target a single distinct attribute

- be unambiguous (phrasing clear and simple)

- be relevant in a verbal face-to-face feedback session

- be generally applicable in scheduled face-to-face feedback discussions

- make sense with the three rating options

Table S7.8 Detailed item revisions, with rationale, to the provisional feedback instrument

Provisional item	Item aim	Any proposed item changes and rationale, with results from Multifaceted Rasch model analysis (MFRM) and exploratory factor analysis (EFA).	Revised item
Item 1: Based on observed performance The educator's comments were based on observed performance	Link the educator's performance comments with specific example of relevant performance, so the learner can understand the basis.	• Combine items 1 & 17 into a single item, as they overlap. • Remove 'based' as it may not be observable. For example, it could occur within the educator's mind or could be satisfied by the context alone (such as workplace-based assessment). • EFA: high loading [.807] with 'analyse performance' alone. • MFRM: no weighted misfit.	*Specific instance* The educator linked their comments to a neutral description of what the learner did (action, decision or behaviour) and the consequences.
Item 2: Timely feedback The educator offered to discuss the performance as soon as practicable	Timely feedback.	• Deleted from the instrument during pilot testing, as it would occur before a feedback conversation.	Deleted
Item 3: Feedback purpose clear The educator explained that the purpose of feedback is to help the learner improve their performance	Explicitly affirm the objective for the feedback conversation, to focus both participants	• Revise concept catcher, to make phrasing simpler and more similar to other concept catchers by using an active verb first. • EFA: high loading [.640] with 'set the scene' alone • MFRM: no weighted misfit.	*Clarify purpose* The educator explained that the purpose of feedback is to help the learner improve their performance.
Item 4: Establish a non-judgmental atmosphere: 'here to help' The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise	Foster a 'learning orientation' that focuses on continually developing skills. Mistakes or skill gaps are expected while learning a skill and the educator's role is to assist the learner to develop their skills, not criticise. This aligns with Dweck's 'growth mindset'.[157]	• Revise the concept catcher. Remove 'non-judgemental', as judgement is an integral component of performance analysis, especially during summative assessments. New concept catcher highlights that mistakes are an expected part of learning and provide learning opportunities. • In the item, more directly address learners' fear regarding negative consequences if performance mistakes or omissions are exposed and highlight the associated learning opportunities. • EFA: high loading [.734] with 'set the scene' alone • MFRM: no weighted misfit.	*Expect mistakes and regard them as learning opportunities* The educator conveyed the view that mistakes or omissions are expected while developing skills and are opportunities to continue refining skills.

Item 5: Clarify feedback process, so learner knows what to expect The educator described the intended process for the feedback discussion	Create an opportunity for the learner to understand and contribute to expectations. The purpose is for the educator to increase transparency and share power over the session plans, in order to reduce excessive learner anxiety related to 'not knowing what to expect', which may interfere with attention, complex thinking and memory.	• Clarify item meaning by changing 'process' to 'expectations' and list key topics to consider. • Clarify the expectation to consider the learner's input regarding the session plan. • EFA: does not load substantially onto any factor. Highest loading is with 'set the scene' [229] but very low communality [.114]. • MFRM: weighted misfit in original analysis (MNSQ =2.2, T =3.0) and sensitivity analysis (MNSQ =2.0, T =2.7) • Behaviour rarely seen: item rated as 0 = 'not done' in 148/174 (85%) sets of video analysis • Phrasing vague.	*Discuss session plan, so learner knows what to expect* The educator discussed expectations for the feedback session, taking the learner's perspective into account. (This may cover the session outline; time available; value of interactive dialogue and collaboration).
Item 6: Encourage dialogue The educator encouraged the learner to engage in interactive discussions	• Promote interactive dialogue, with both participants contributing and building on each other's comments, for the purpose of establishing a shared understanding and co-constructing ideas.	• Make item more explicit regarding what constitutes 'interactive dialogue' and add supplementary list of illustrative examples, based on parallel qualitative research on psychological safety.[242] • EFA: high loading [.797] with 'foster learner agency' alone • MFRM: weighted misfit in original analysis (MNSQ =2.3, T = 4.0) and sensitivity analyses, but less (MNSQ = 1.7, T = 2.6) • Potential contributory factors: -Phrasing vague.	*Promote interactive dialogue* The educator encouraged discussions in which both perspectives were shared and thoughtfully responded to. (This may include agreeing, seeking more information, checking information, asking questions, building on an idea or contesting an idea with the rationale).
Item 7: Seek learner's priorities The educator asked the learner about their learning priorities for the observation and feedback discussion, and responded to them	• Find out about the learner's learning priorities for the session.	• In the item, remove 'observation' as this instrument concerns the feedback discussion, not the prior observation of performance. • Use 'explore' instead of 'ask', to emphasise that a single question may be insufficient, and the educator may need to make additional efforts to discover the learner's priorities. This phrasing also makes sense with each rating category. • Revise phrasing, so the meaning is clearer and simpler. • Keep item within 'set the scene' as it is important to find out a learner's priorities at the start.	*Seek learner's priorities* The educator explored what the learner most wanted to raise in the feedback session.

		• EFA: moderate loading with two factors: 'set the scene' [.566] and 'foster learner agency' [.465]. This makes sense as this item contributes to both. • MFRM: no weighted misfit. • Middle rating category (1= done somewhat or sometimes) does not make sense with 'ask', which results in a dichotomous outcome (educator did/did not ask).	
Item 8: Encourage learner to 'work it out for themselves' The educator encouraged the learner to consider the issues and possible solutions during the feedback discussion.	• Promote active learning by supporting learner to reflect and problem solve, aligned to a social constructivist paradigm.	• In the concept catcher, change 'encourage' to 'support', to reduce duplication. • EFA: high loading [.770] with 'foster learner agency' alone • MFRM: weighted misfit in original analysis (MNSQ =1.7, T =2.6) but not the sensitivity analysis (MNSQ =1.5, T =1.9).	*Support learner to 'work it out for themselves'* The educator encouraged the learner to consider the issues and possible solutions during the feedback discussion.
Item 9: Encourage learner to focus on learning, rather than trying to cover up limitations The educator encouraged the learner to discuss difficulties and ask questions regarding the performance so the educator could help the learner to develop solutions.	• Deliberate focus on encouraging learning behaviours, in particular 'what was difficult?' and 'any questions?' , in order to offer assistance, in line with encouraging a 'Growth mindset'.[157]	• No change. • EFA: high loading [.709] with 'foster learner agency' alone • MFRM: no weighted misfit.	No change.
Item 10: Acknowledge learner's emotional response The educator acknowledged and responded appropriately to emotions expressed by the learner.	• Respond appropriately to learner's emotional reaction during a feedback discussion. Emotions are commonly stimulated and if not attended to, they can consume attention and hence hinder reasoning, learning and memory.	• In the concept catcher, use 'attend' (i.e. 'pay attention to'), as this best captures the intention of the item (whereas the item must be strictly observable). • In the item, remove 'acknowledged', as this is included within 'respond'. • EFA: In a 5-factor solution with oblimin rotation, high loading [.735] with 'foster psychological safety'; Communality .68 • MFRM: no weighted misfit	*Attend to learner's emotions* The educator responded appropriately to emotions expressed by the learner.

Item 11: 'Best interests at heart' The educator showed respect and support for the learner.	• Demonstrate respect and support for the learner, with the aim of working with the learner to foster a psychologically safe learning environment that encourages learning behaviours and intrinsic motivation (which requires sentiments of autonomy, relatedness and competence, in line with self-determination theory).	• Replace this with new additional items that describe specific educator behaviours to demonstrate respect and support. • Item is broad and does not describe specific observable behaviours. • EFA: In a 5-factor solution with oblimin rotation, high loading [.607] with 'foster psychological safety'; Communality .73 • MFRM: no weighted misfit.	Replaced by new additional items.
Item 12: Clarify the value of self-assessment The educator asked what the learner understood about the benefits of self-assessment and helped clarify.	• Promote the value of a thorough learner self-assessment: to practise evaluative judgement and to reveal their understanding of the target performance and provide an opportunity for learner to proffer their learning needs.	• In the item and concept catcher, use 'promoted'' instead of 'ask, as 'ask' can be 'one off' and superficial. This also makes sense with each rating category. • Add supplementary information to explain the value of self-assessment, to assist the educator, informed by the qualitative analysis on evaluative judgement. • EFA: moderately high loading [.519] with 'foster learner agency' alone • MFRM: no weighted misfit • Middle rating category (1= done somewhat or sometimes) does not make sense with 'ask', which results in a dichotomous outcome (educator did/did not ask). • Behaviour rarely seen: item rated as 'not done' in 133/173 (77%) sets of video analysis data.	*Promote the value of self-assessment* The educator promoted the benefits of self-assessment in discussion with the learner. (Self-assessment provides opportunity for a learner to describe key features of the target performance, analyse their own performance in comparison of the target performance and raise learning needs; this involves practising valuable learning skills such as reflection, evaluative judgement and self-regulated learning).
Item 13: Learner self-assessment The educator asked the learner to identify key similarities and	• Thorough self-assessment by the learner, involving a comparison of the learner's performance with the target performance.	• In the item, use 'encouraged' instead of 'asked', to emphasise that a single question is often insufficient, in which case the educator may need to make additional efforts to enable the learner to voice their self-	*Learner self-assessment* The educator encouraged the learner to identify key similarities and differences between the

differences between the learner's performance and the target performance.		assessment, informed by the qualitative analysis on evaluative judgement. This also makes sense with each rating category. • Position with 'Analyse performance' to emphasise inclusion of the learner's self- assessment. • EFA: moderately high loading with 'foster learner agency' [.522] and lower loading with 'analyse performance' [.358] • MFRM: no weighted misfit • Middle rating category (1= done somewhat or sometimes) does not make sense with 'ask', which results in a dichotomous outcome (educator did/did not ask).	learner's performance and the target performance.
Item 14: Target performance and reasoning clear The educator clarified with the learner key features of the target performance and explained the reasoning.	• Explain what the task 'should look like' and the rationale, to strengthen the learner's understanding of the target performance, particularly in a specific context	• No change, informed by the qualitative analysis on evaluative judgement. • EFA: high loading within performance analysis alone [.777] • MFRM: mild weighted misfit in original analysis (MNSQ =1.7, T =2.3) and sensitivity analysis (MNSQ =1.7, T =2.2) No problems apparent on review.	No change
Item 15: Educator assessment, including clear performance gap The educator clarified with the learner similarities and differences between the learner's performance and the target performance.	• Comparison of the learner's performance with the target performance, to clarify i) which aspects were done effectively, thereby promoting feelings of competence and reward for the effort put into skill development; and ii) which aspects were not done effectively, thereby clarifying performance gap, which focuses attention and motivates change. • Typically, this is most useful when focused on the learner's zone of proximal development.[22]	• In the concept catcher, use 'analysis' instead of 'assessment', to focus on the comparison of the learner's performance with the target performance, in line with a learning orientation, as opposed to 'assessment', which could imply just a 'score' or a judgement that the performance was 'good enough' (or not). • Add supplementary recommendation to consider focusing on the zone of proximal development, to enhance developmental focus. • EFA: high loading [.768] with 'analyse performance' alone • MFRM: minor weighted misfit in original analysis (MNSQ =1.6, T =2.1) but not in the sensitivity analysis (MNSQ =1.2, T =0.9).	*Educator's performance analysis* The educator clarified with the learner similarities and differences between the learner's performance and the target performance. (Consider focusing on the 'development zone' i.e. around their current position on the learning curve).

Item 16: Educator comments on a few, important issues The educator's comments focused on key issues for improving the performance	• Prioritise a few comments and avoid numerous or low value comments which contribute to cognitive overload. Select topics which are important and most likely to improve performance.	• Shorten the concept catcher, to make phrasing simpler and add 'prioritise' for emphasis. • In the item, add 'a few' as an important observable feature and change 'issues' to points' as 'issues' may imply 'problems'. • Add a supplementary recommendation to consider improvements beyond just the current task, to assist the educator. • EFA: high loading [.614] with 'Analyse performance' alone. • MFRM: weighted misfit in original analysis (MNSQ =1.95, T =3.1) but not in the sensitivity analysis (MNSQ =1.4, T =1.3).	*Prioritise comments* The educator's comments focused on a few, key points for improving performance. (This may focus on improving the current task or broader learning strategies).
Item 17:Specific instance ('what happened') First the educator described, using neutral language, what the learner did (action, decision or behaviour), and the consequences	• Items 17-19 were originally designed to address corrective comments (only). • Link educator's comments with specific example/s of relevant performance, so the learner can understand the basis. • Use neutral language when describing the performance (i.e. avoid critical language), to isolate 'observed performance' from the educator's interpretation or evaluation.	• Combine items 1 & 17 into a single item, as they overlap. • Make item 'generally applicable (items 17-19 were originally designed to address corrective comments only). • EFA: high loading [.705] with 'analyse performance' alone MFRM: no weighted misfit.	See revised Item 1
Item 18: Educator's perspective clear ('why it matters') The educator clearly explained their perspective on the learner's actions, including the reason for their concern	• Items 17-19 were originally designed to address corrective comments (only). • Clear explanation of the educator's comments about the learner's performance (information or their opinion), accompanied by the rationale. This assists the learner's understanding of the educator's perspective. It also supports transparency and autonomy by providing substantiation, to allow	• Broaden the item to apply whenever the educator provides information or their opinion, so it is 'generally applicable'. Items 17-19 were originally designed to address corrective comments (only). • EFA: high loading [.665] with 'Analyse performance' alone. • MFRM: no weighted misfit.	*Educator's perspective clear* The educator's comments (information or opinion), accompanied by the reasoning, were clearly explained.

	the learner to judge the validity for themselves		
Item 19: Educator explores learner's perspective ('why' learner acted as they did) The educator explored the learner's perspective and reasoning to reveal the basis for the learner's actions (e.g. what was the learner trying to do and options considered/ difficulties encountered)	• Items 17-19 were originally designed to address corrective comments (only). • Important for the educator to explore and understand the learner's perspective, particularly the drivers for the learner's actions, so the correct learning needs can be targeted	• Broaden the item, to highlight the value of exploring and listening attentively to the learner's perspective at all times, so the item is 'generally applicable'. This shows respect and humility by valuing the learner's perspective (showing interest & willing to be influenced by it). • Position item within 'Foster psychological safety' as it fits best there, once generalised. • EFA: high [.635] loading with 'Foster learner agency'. • MFRM: no weighted misfit.	*Value learner's perspective* The educator explored the learner's perspective and reasoning, and demonstrated attentive listening.
Item 20: Focus on actions, not the person ('did' not 'is') The educator's comments were focused on the learner's actions not personal characteristics	• Focus the learner's attention on improving task performance and avoiding attention on 'self', which risks strong emotional reactions concerning innate characteristics.[2]	• No change • EFA: high loading [.798] with 'Analyse performance' alone. • MFRM: no weighted misfit.	No change
Item 21: Select learning priorities: most useful (important and relevant) for the learner The educator helped the learner to select a couple of key aspects of the performance to improve	• Educator and learner deliberately select 1-2 learning goals that are most useful for the learner, in order to improve performance. The learner's involvement is important to support autonomy and motivation.	• In the item, remove 'the' from 'the performance' to expand the focus to include both the current task and broader learning strategies. • Change 'a couple of' to 'few' for to make the phrasing simpler. • Add supplementary information to recommend potentially valuable options. • EFA: moderately high loading [.523] with 'Plan improvements' alone • MFRM: no weighted misfit.	*Select learning priorities: most useful (important and relevant) for the learner* The educator helped the learner to select a few, key aspects of performance to improve (Consider learner's request; key mistake or omission or 'next steps').
Item 22: Develop the action plan: how to do it! The educator helped the learner to work out how they could improve their performance and	• Educator and learner develop specific and practical strategies to improve performance.	• Revise concept catcher and item, to make phrasing simpler. • Expand improvement focus to include both the current task and broader learning strategies. • EFA: moderately high loading with [.553] 'Plan improvements' alone • MFRM: no weighted misfit.	*Develop the action plan* The educator helped the learner to develop specific and practical plans to improve their performance.

specify the practical steps to achieve it.			(The plans may relate to the current task and/or broader learning strategies).
Item 23: Check the learner understands the plans The educator checked if the learner understood their learning goals and action plan by asking them to summarise it in their own words	• Check the learner's understanding of the learning goals and action plan, using a teach-back technique.[270]	• Combine items 23 & 24 to reduce overlap. • Shorten the concept catcher, to make phrasing simpler. • EFA: high loading [.727] with 'Plan improvements' alone. • MFRM: mild weighted misfit in original analysis (MNSQ =1.8, T =2.4) and less in sensitivity analysis (MNSQ =1.7, T =2.1). • Behaviour rarely seen: rated as 'not done' in 139/173 (80%) sets of video analysis.	*Check plans and rationale understood* The educator checked if the learner understood their learning goals and action plan, accompanied by the rationale, by asking them to summarise it in their own words.
Item 24: Checks the learner understands the rationale: 'why it's better' The educator checked if the learner understood the rationale for their learning goals and action plan	• Check the learner's understanding of the rationale underpinning the learning goals and action plan, using a teach-back technique.[270]	• Combine items 23 & 24 to reduce overlap. • EFA: high loading [.602] with 'Plan improvements' alone. • Rasch: no weighted misfit.	Combined with item above
Item 25: Plan opportunities to review the impact of the feedback The educator discussed with the learner possible subsequent opportunities for the learner to review their progress	• Plan how to review progress and whether the action plan resulted in improved performance, to complete the feedback loop. Subsequent performance may be reviewed by another clinician.	• Shorten the concept catcher, to make phrasing simpler. • Remove 'subsequent' as unnecessary, to make phrasing simpler. • EFA: low loading [.474] with 'Plan improvements' and communality low at .237. • MFRM: no weighted misfit. • Behaviour very rarely seen: item rated as 'not done' in 165/172 (96%) sets of video analysis data.	*Plan how to review progress* The educator discussed with the learner possible opportunities for the learner to review their progress.

New items to expand domain 'foster psychological safety' based on the qualitative analysis[242]			
New item	• Educator demonstrate respect, humility and inclusiveness by appreciating the learner's contributions, attributes, skills and potential.		*Appreciate learner's contributions* The educator expressed appreciation for the learner's contributions. (This may include the learner's input into the discussion; learner's contribution to healthcare practice; learner's attributes, skills, or future potential).
New item	• Educator demonstrate respect for learner's autonomy		*Respect learner's autonomy* The educator showed respect for the learner's autonomy. (This may include encouraging the learner to take a turn to lead the conversation; state their opinion or preference; make a choice; or contest the educator's comments for the purpose of learning).
New item	• Educator demonstrate support and care by expressing compassion for difficulties experienced by the learner.		*Show compassion* The educator expressed compassion for difficulties experienced during training, raised by the learner
New item	• Educator demonstrate humility by recognising own limitations		*Show humility and recognise own limitations* The educator conveyed the view that everyone has limitations, including themselves.

			(This may include acknowledging educator's limitations e.g. routine uncertainty during clinical practice; beyond their speciality; educator's evaluation, opinion or advice contestable; made mistakes themselves while learning; or general limitations e.g. 'always more to learn' or 'a common mistake').

7.9 Chapter 7 Closing discussion

The article in Chapter 7 described the completion of Phase 3, Stage 2 containing the psychometric analysis of the provisional instrument; and Phase 3, Stage 3 involving the refinement of the provisional instrument and presentation of the Feedback Quality Instrument.

The publication only briefly touched on how the thematic analyses on psychological safety and evaluative judgement informed the way that related items were shaped in the revised instrument, summarised in the supplementary information: Table S7.8. The EFA identified items related to psychological safety as a core construct within quality feedback, resulting in a domain called *Fostering psychological safety* within the FQI. The EFA also indicated the need for additional items to expand this domain, to create a richer portrayal of practical ways educators might do this. Some examples of how the study findings informed the development of new items were described in the closing discussion of Chapter 6. As promoting psychological safety is an activity and goal throughout the feedback interaction, the findings from the qualitative analysis also led to revisions of previous items, as well as the creation of new ones. As one example, as described in the closing discussion of Chapter 3, it had been difficult to reach consensus on Item 4 in the provisional instrument. The key considerations concerned how to 1) have a 'non-judgemental attitude' in the context of a performance analysis and 2) 'accept mistakes' in clinical environments where high standards are required to keep patients safe. The theme, called a '*continuing improvement orientation*' from the qualitative analysis on psychological safety, detailed in Chapter 6, played a key role here. Dweck called this a growth mindset.[157] The theme highlighted how some educators supported learners to continually enhance their skills by normalising mistakes, assisting learners to correct them and rewarding progress. These practical strategies seemed to alleviate learners' fear of unrealistically high performance expectations. As a consequence, the item in the provisional instrument:

Establish a non-judgmental atmosphere: 'here to help'

The educator indicated that while developing a skill, it is expected that some aspects can be improved and the educator is here to help, not criticise.

was rephrased to:

Expect mistakes and regard as learning opportunities

The educator conveyed the view that mistakes or omissions are expected while developing skills and are opportunities to continue refining skills.

Concerning evaluative judgement, some examples of how the qualitative study findings informed item revisions were outlined in the closing discussion of Chapter 5.

The key concepts, or pillars supporting quality feedback, discovered in EFA led to a potentially more useful clinical version of the Feedback Quality Instrument as it provided a pathway for grouping like items into five domains. A clinician is likely to find it easier to focus on five overarching concepts when they reflect on their practice, in comparison to consideration of 25 individual items.

This article concludes the output from the PhD research program and accomplishes the objective to develop an instrument that clearly describes practical ways that educators can collaborate with learners to optimise learner benefits. Nevertheless, this is more of a 'staging post' than 'the finish'. While recognising the limitations of the current Feedback Quality Instrument, it has many useful applications in its current form and presents a solid platform for future growth and improvement.

Chapter 8

Synthesis of Study Findings and Conclusion

8.1 Introduction

This final chapter presents an overarching summary of this PhD program of research that employed a multiphase mixed methods design. The aim of the PhD was to develop a Feedback Quality Instrument, to analyse and guide educators in fostering quality face-to-face feedback interactions with learners in clinical practice. When this work commenced there were many important gaps in knowledge that are now much clearer. The chapter includes a summary of the research findings that address each research question, the implications of those findings, potential future research directions, research strengths and weaknesses, and a conclusion.

8.2 Summary of research findings

Phase 1: Systematic review

RQ 1: What is the current evidence available regarding the impact of face-to-face feedback on workplace task performance involving health professionals?

This question targeted the key premise underlying the entire research program, that is, can feedback assist health professionals to improve their performance? In Phase 1, described in Chapter 2, the systematic review summarised and analysed the available evidence on the impact of face-to-face verbal feedback on objective task performance of health professionals, compared to no or alternative feedback. The first comparison of face-to-face feedback with no feedback included eight studies with 392 health professionals, when studies at high risk of bias were excluded. Meta-analysis showed a moderately large effect size of 0.7 (95% CI 0.37 to 1.03; $p<0.001$) in favour of feedback and the calculated SMD prediction interval was -0.06 to 1.46. For the first time, this provided empirical support for the assertion that face-to-face feedback enhances health professionals' performance. The wide range in effect suggested that specific feedback components may be more beneficial than others. The second comparison of face-to-

face feedback with alternative feedback revealed a wide variety of feedback interventions but meta-analysis was not possible due to heterogeneity in the studies. Overall, the quality of evidence was found to be low, primarily due to risk of bias in study design and publication bias. In addition, most of the studies were small and one off, and did not clearly describe all the important aspects of their protocols, including feedback components. This indicated the need for a systematic approach to investigating the effect of different components within feedback and established the potential value of developing a Feedback Quality Instrument. Such an instrument could bring together the key components in feedback that enhance learner performance and promote effective feedback. It could also provide the foundation for systematically evaluating which components of feedback or their combinations have the greatest influence on leaner performance.

Phase 2: Creating a provisional feedback instrument

Phase 2 resulted in the creation of a provisional feedback instrument. The two discrete stages in Phase 2 were described in Chapter 3.

RQ 2: What are the key educator roles in promoting quality feedback discussions with learners that result in beneficial learner outcomes?

In Phase 2, Stage 1, an additional extensive literature review, utilising a snowball technique, searched for 1) discrete elements of an educator's role during face-to-face feedback interactions considered to influence learner outcomes and what research supported each element and 2) previously published face-to-face verbal feedback assessment instruments used in health professions education. Over 170 relevant articles were identified across diverse literature and extracted information was used to delineate eighteen discrete elements of an educator's role in quality feedback. Ten face-to-face verbal feedback instruments were identified but none were designed to analyse an educator's role in a face-to-face feedback discussion concerning a learner's clinical performance.

RQ 3: Could consensus be achieved on a comprehensive set of statements that describe important educator behaviours in quality feedback?

In Phase 2, Stage 2, these elements were operationalised into an initial list of twenty-three descriptions of corresponding observable educator behaviours by the core research team. Following three rounds of a Delphi process plus a final face-to-face meeting with a panel of eleven experts, consensus was achieved on a comprehensive set of twenty-five descriptions of important educator behaviours in quality feedback.

Phase 2 established the components of quality learner-centred feedback interactions for educators, supported by research information and endorsed by experts. The set of observable educator behaviours constituted a provisional instrument with 25 items.

Phase 3: Testing, analysing and refining the provisional instrument, to create the Feedback Quality Instrument

Phase 3 had 3 Stages that described testing and analysing the provisional instrument using both qualitative and quantitative methods that informed refinements and the development of the Feedback Quality Instrument.

RQ 4: What can be learnt about current feedback practices, by analysing data gained through administering the provisional instrument to systematically analyse feedback interactions in clinical practice?

Phase 3, Stage 1, described in Chapter 4, started by collecting 36 video-recordings of authentic feedback interactions during routine clinical practice. The participants were 34 educators, including 26 medical, 4 nursing and 4 physiotherapy professionals with a mixture of specialities, supervisor experience and gender; and 35 learners from the same health professions, also with a mixture of experience and gender. This provided rare observational data involving diverse health professionals in hospital practice. Next, a team of raters independently administered the provisional instrument to rate educators' feedback practice seen in the videos, compared with recommended educator behaviours described in the provisional instrument, to produce 174 sets of ratings data.

Each behaviour was rated 0 = not seen, 1 = done somewhat, 2 = done. The raters also participated in usability analysis, reported in Chapter 7.

The data from this systematic analysis was used to compile a descriptive report on contemporary clinical educators' feedback practice. It revealed marked variation in how frequently each recommended behaviour was demonstrated, based on the mean score across all the feedback videos. Individual educators' practice differed also, based on the sum of rating scores for each of the recommended 24 behaviours, as one behaviour from the original 25 was excluded. This report provided a compelling argument on the need for clear guidance for clinical educators on the features of quality feedback. In addition, it shed light on which components of quality feedback had, or had not yet, been translated from the literature into routine practice by everyday clinicians. This provides the groundwork and drive for designing strategies to assist educators and targeting professional development training to employ more effective feedback behaviours.

RQ 5: What novel insights can be gained from thematic analysis of feedback episodes in clinical practice regarding emerging and under-researched arguments regarding best practice in feedback using the examples of i) learner evaluative judgement and ii) learner psychological safety?

Phase 3, Stage 2 centred on analysing data generated by testing the provisional instrument in parallel studies using both qualitative and quantitative methods. Qualitative analysis of the feedback video transcripts was used to explore two emerging aspects of quality feedback: evaluative judgement, described in Chapter 5 and psychological safety, described in Chapter 6. In the study into evaluative judgement, the themes developed were 1) *Educator invitations: I want to know what you think about the quality of your work'*; 2) *Good work looks like this* and 3) *Calibration of the learner's judgement through dialogue*. These findings contributed to refining items that related to educators' probes to encourage learners to express their own judgements about the quality of their own work and explicit discussion of what constitutes good work (explicating standards).

The study into psychological safety highlighted its influential role in learning and performance, as learners take a risk when they engage in productive learning behaviours. The findings enhanced understanding around 'what psychological safety looks like' in workplace feedback. Four themes were developed were 1) *Setting the scene for dialogue and candour*; 2) *Educator as ally*; 3) *A continuing improvement orientation*; and 4) *Encouraging an interactive dialogue*. Each revealed practical approaches that educators used to promote a psychologically safe learning environment together with learners.

These qualitative analyses clarified how theoretical principles recommended in the literature could be manifest in feedback discussions. The observable behaviours identified were incorporated into items in the Feedback Quality Instrument, to assist educators to improve opportunities during feedback to foster psychological safety with learners and to involve learners in developing their own evaluative judgement.

RQ 6: How can a feedback quality instrument be refined and made ready-for-use, based on multiple inputs including i) usability testing ii) psychometric analysis involving multifaceted Rasch model analysis and exploratory factor analysis iii) thematic analysis in under-researched areas?

The last article, in Chapter 7, described the parallel quantitative analysis, involving psychometric analysis, of data generated by testing the provisional instrument to complete Phase 3, Stage 2 and then the process of refining the provisional instrument to present the Feedback Quality Instrument (FQI), ready-for-use, in Phase 3, Stage 3.

Usability analysis by the raters, performed around administration of the provisional instrument (detailed in Chapter 4), identified difficulties with specific items, the rating scale, plus the demand incurred by rating many items. Psychometric analysis involved exploratory factor analysis and multifaceted Rasch model analysis. Exploratory factor analysis identified clusters of highly correlated items, indicating five core concepts underlying quality feedback, for educators to: foster psychological safety, foster learner agency, set the scene, analyse performance and plan improvements. It also identified

that there were insufficient items characterising psychological safety; this triggered the process to design additional items, which prompted the qualitative study on psychological safety, described in Chapter 6. The Rasch analysis provided information on specific aspects of the instrument, items and rating scale that were not usefully contributing to accurately estimating educators' feedback proficiency.

Revisions to the provisional feedback instrument were informed by multiple inputs including usability testing; psychometric analysis involving multifaceted Rasch and exploratory factor analysis; parallel research into psychological safety; and underlying research and theory. This resulted in the Feedback Quality Instrument with 25 items that describe key observable educator behaviours considered to engage, motivate and enable a learner to improve. The items are grouped according to the five key concepts underlying quality feedback, identified in the exploratory factor analysis: set the scene, analyse performance, plan improvement, foster learner agency, and foster psychological safety. The FQI, based on a meticulous research process, provides practical descriptions of educators' behaviours designed to promote high quality feedback discussions with learners.

8.3 Unique contributions of this PhD

This PhD has made the following unique contributions to promoting quality feedback interactions in the health professions:

- The systematic review and meta-analysis provided evidence to support the assertion that verbal feedback is likely to substantially enhance health professionals' workplace performance. This highlights the value of attempting to optimise feedback interactions.
- The development of a comprehensive set of statements describing key educator behaviours considered to enhance learner outcomes, supported by research and endorsed by experts
- Video-recordings of authentic feedback interactions involving diverse health professionals provided rare and rich insights. These data enabled a characterisation of contemporary feedback discussions that revealed gaps in translation from theory into practice, particularly the lack of learner involvement

in the process. It also revealed how theoretical principles regarding psychological safety and evaluative judgement recommended in the literature could be manifest in feedback interactions.

- A Feedback Quality Instrument that clearly articulates specific ways for educators to collaborate with learners in quality, learner-centred feedback discussions, was developed using an extensive 3-phased research process. This offers practical advice for clinical educators and a foundation for future research into systematically evaluating the influence of different constituents of educators' feedback practice on learning and performance.

8.4 Implications

This thesis, in developing the Feedback Quality Instrument, has clarified a suite of practical ways for educators to support quality learner-centred feedback in clinical practice, substantiated by empirical research. The FQI applies to educators' practice but the goal is to support learners. The domains highlight five major concepts underlying quality feedback and the specific descriptions of helpful behaviours in quality feedback, within the FQI, provide educators with clear explanations on how to translate important feedback principles into practice.

The FQI presents many valuable opportunities to enhance both educator and learner feedback literacy and evaluative judgment within the health professions. Starting with educators, they could use the FQI as a framework to reflect on their own feedback practice, by identifying behaviours that they already do routinely, thereby confirming good practice and most likely, learn a few new ones that could enhance their approach. They could simply review the instrument in preparation for a feedback encounter as a sensitising technique, maybe focusing on one or two behaviours that they wanted to embed in the upcoming interaction. Taking advantage of the benefits of peer learning, a colleague could be invited to observe the feedback interaction, if other participants agreed, to offer an outsider's perspective and work with the educator to enhance their feedback skills. Alternatively, again with the consent of all involved, an educator could video the feedback episode and review it alone or with others, in confidence. As

mentioned earlier in this discussion, all of these possibilities could be enhanced by involving learners. However, as discovered in this PhD, many health professionals feel reticent to have feedback practice scrutinised so it may help to directly acknowledge this. Moreover, additional safeguards could be implemented, particularly for videos or other enduring artifacts that could be shared, such as putting in place such a more formal consent with a strict confidentiality clause and a requirement to delete the video once reviewed a short while later, as occurred in our research.

In professional development initiatives, the FQI could be used to discuss key characteristics of quality feedback, looking at it from both the educators' and learners' perspectives. It would help educators and learners to enhance their understanding of 'good feedback practice' to consider both their own and the other's role. In particular, highlighting those items that were rarely seen in the research videos, has the potential to stimulate important changes in feedback practice. While watching role play or videos of feedback discussions, the FQI could be used by educators and learners to scrutinise interactions, match moments with corresponding items, select items they most wanted to discuss or to suggest improvements to observed practice. Another option could be to involve educators and learners in 'workshopping' scenarios to rehearse what educators might say, using their own words in their own style, to 'get comfortable' with how they could do this in real life. Involving learners would allow educators to find out how their actions were being perceived and would allow for better tracing of the effect of some of these actions. For learners, it would offer insights into the educator's role, and how learners themselves could contribute to support quality feedback interactions. Examples might include learners taking on an active role by voicing a genuine self-assessment or asking educators to discuss what they should prioritise to improve.

The feedback videos have provided rich data, showcasing different capacities of educators to support productive dialogue. This may be used in the form of case studies to help learners and educators come to understand typical patterns of engagement in feedback to encourage a pivot to more productive habits. Examples of good practice may also be used in professional development initiatives to help both parties to better

understand how items can be manifest in practice. In the near future, it is planned to create educational videos by extracting key exemplars of high and lower quality practice that can be de-identified and re-enacted. Using authentic exemplars in such training videos is likely to provide helpful teaching opportunities.

8.5 Strengths and limitations

The strengths in this PhD research program arise from the rigorous mixed-methods design and richness of the videos of authentic feedback interactions. A systematic review verified the main premise that feedback has the potential to substantially enhance health professionals' performance. A provisional instrument was created by identifying discrete components of feedback substantiated by published research then gaining expert consensus on a set of observable educator behaviours, comprising a provisional instrument, using a Delphi technique. A large set of videos of authentic feedback interactions with diverse health professionals were collected. Educators' practice in the videos was rated using the provisional instrument. These ratings were used to create an observational report that clarified current feedback practice. Quantitative psychometric analysis of the ratings data, using exploratory factor analysis and multifaceted Rasch analysis, and qualitative analysis of the video data, leading to novel insights on two key under-researched areas, were conducted. Finally the Feedback Quality Instrument was created, informed by usability analysis, quantitative and qualitative analyses, and foundational theory. To our knowledge, no other feedback instrument in clinical practice has undergone such meticulous research, originating from published literature and expert collaboration, then refined based on both quantitative and qualitative analyses.

The FQI has the potential to appreciably change feedback in clinical practice, arising from the practical assistance offered by the explicit descriptions of educator behaviours considered to foster quality learner-centred feedback, and its potential use in future research, particularly to investigate which components have the greatest impact on learner outcomes.

However, this research has limitations. All the videos were recorded at one health service, so this may limit the generalisability of the findings, although the health service is very large and employs health professionals from many regions of the world. Participants volunteered for the study, so the group may have been more confident in their feedback skills and not fully reflect the typical range in practice. Recording feedback interactions could have influenced behaviour and it may be that participants 'put on their best performance'. Again, this would affect the findings of this research, although a number of videos showed participants commenting that they had forgotten it was recording, when they turned their attention to switching the recording device off. There were some problems associated with administering the instrument, including interpreting items and applying them consistently, but it is anticipated that this will be improved by the revisions and introduction of a handbook. In addition, there were differences in rater severity. This may be improved following training and calibration, but it is accounted for during the multifaceted Rasch analysis. The qualitative analyses relied on observing participants' behaviour and then making inferences, without any direct interview data to understand participants' experiences or perspectives. Hence interpretations may have been incorrect, however looking for consistency across videos and involving multiple researchers, attempted to limit this. Both the qualitative and quantitative analyses involved 'judgement' to interpret and apply the findings to developing the FQI, so it is important to acknowledge the influence of the primary researcher, in particular.

8.6 Future research

There are many opportunities to extend the research outlined in this thesis. The literature search for the systematic review was repeated intermittently across the period of this PhD. Each time, more and higher quality studies were found. Therefore, repeating the systematic review and meta-analysis in the future could provide more robust evidence regarding the effects of face-to-face feedback on health professionals' performance.

Currently neither educators nor learners have been actively involved in the FQI's development but their contributions will be essential to further refinements and then implementation. The next phase in developing the Feedback Quality Instrument will involve further usability testing including 'thinking out loud' testing,[256] and testing the FQI with larger numbers, to enable more robust psychometric analysis. This will also provide the chance to evaluate revised items and the new items on psychological safety, and refine them as required. Multifaceted Rasch model analysis will enable the development of a linear scale for measuring educator proficiency in quality feedback. This will aid accurate comparisons in future research investigating the influence of different educator behaviours on learner outcomes. It will also help educators to monitor their progress in developing feedback skills. The Rasch analysis, by ranking items according to difficulty, could enable professional development in feedback to guide educators' progress along the expected learning curve. Once finalised, the FQI will provide the foundation for a research program to systematically investigate the effect of different educator behaviours on learner outcomes and clarify their relative value. It may also be helpful to provide illustrative examples of what an educator may say or do in achieving the items (using research data, or reconstituted case studies informed by this research).

It will be important to see how the FQI works in practice, to clarify the most effective ways to use it. The FQI has 25 items and offers a comprehensive view of 'what quality feedback for educators looks like'. Grouping items into five domains, representing core concepts in quality feedback, may make it easier for users. In addition, developing a brief and user-friendly manual would be helpful, to provide additional information about the FQI instrument, items, ratings and scores. However, the large number of items may still create a substantial cognitive load. This could be reduced further by producing a digital application (or 'app') that could provide prompts to guide the user and could also incorporate the manual and scoring. Nevertheless, everyday clinicians, wanting to improve their feedback practice, are likely to prefer a shorter version; in this situation *simplicity* may be a higher priority than *comprehensive*. Therefore, an abridged version, incorporating just the most influential behaviours could substantially increase

engagement and application. Then, for those who wished to advance their skills further, they could graduate on to the more detailed, 25 item version.

Future research could explore the benefits or obstacles for educators and learners associated with using the FQI in different ways. A few examples include educators reviewing the FQI before or after a feedback interaction, either alone or with a peer; finding workable ways to support changing practice during feedback interactions, maybe using a coach or cue cards (or an e-version) based on the FQI; or focusing on the educator alone or bringing in the learner, for example by sharing their perspective on the feedback session afterwards, to incorporate both sides of the dynamic.

It may be possible to use the FQI to explore the impact of other influences on feedback outcomes, such as different contexts. Research studies might compare groups across different professions or countries, to answer the question: How does the professional field (nature of clinical practice) influence what happens in the feedback sessions?

This research has focused on the educator's role, representing 'one partner in the dance'. It would be interesting and useful to collaborate with those researching the learner's role. This could provide opportunities to be informed by each other's work and collaborate to develop new insights together, echoing what happens in quality feedback interactions.

8.7 Conclusion

This PhD has described the nature and effects of current feedback in the health professions through a systematic review, and observational study of feedback interactions in the workplace. A series of iterative studies resulted in the development of the Feedback Quality Instrument, with 25 items, now ready-for-use by educators in clinical practice. The FQI contains five domains portraying core concepts that constitute high quality feedback. Three domains occur sequentially, *set the scene, analyse performance* and *plan improvement* and two flow throughout a feedback encounter, *foster psychological safety* and *foster learner agency*. This instrument offers educators

a set of explicit descriptions of useful behaviours to guide workplace feedback and guide educators' evaluative judgement when it comes to feedback. Both educators and learners may benefit from engaging with the instrument to sensitise them to feedback approaches that encourage the learner's engagement, and that promote co-construction of strategies for improvement. The FQI also provides a platform for systematic analysis of the influence of educator behaviours during feedback on learner engagement and subsequent performance.

Bibliography

1. Hattie J, Timperley H. The Power of Feedback. *Review of Educational Research.* 2007;77(1):81-112.
2. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin.* 1996;119(2):254-284.
3. Ende J. Feedback in clinical medical education. *Journal of the American Medical Association.* 1983;250(6):777-781.
4. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher.* 2007;29(9):855-871.
5. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine.* 2008;15.
6. Molloy E, Boud D. Changing conceptions of feedback. In: D. B, Molloy E, eds. *Feedback in higher and professional education.* London: Routledge 2013:11-33.
7. Bing-You RG, Stratos GA. Medical students' needs for feedback from residents during the clinical clerkship year. *Teach Learn Med.* 1995;7.
8. Weller JM, Jones A, Merry AF, Jolly B, Saunders D. Investigation of trainee and specialist reactions to the mini-Clinical evaluation exercise in anaesthesia: implications for implementation. *Br J Anaesth.* 2009;103(4):524-530.
9. Cohen SN, Farrant PBJ, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol.* 2009;161.
10. Dawson P, Henderson M, Mahoney P, et al. What makes for effective feedback: staff and student perspectives. *Assessment & Evaluation in Higher Education.* 2018:1-12.
11. Bing-You RG, Paterson J, Levine MA. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Medical Teacher.* 1997;19(1):40-44.
12. Ende J, Pomerantz A, Erickson F. Preceptors' strategies for correcting residents in an ambulatory care medicine setting: a qualitative analysis. *Academic Medicine.* 1995;70(3):224-229.
13. Lockyer J, Violato C, Fidler H. Likelihood of change: a study assessing surgeon use of multisource feedback data. *Teaching and Learning in Medicine.* 2003;15.
14. Boud D, Molloy E. Rethinking models of feedback for learning: the challenge of design. *Assessment and Evaluation in Higher Education.* 2013;38(6):698-712.
15. Physicians RACo. Supervisor Professional Development Program Workshop 1: Developing trainee expertise and using coaching techniques to improve feedback delivery https://www.racp.edu.au/fellows/supervision/supervisor-professional-development-program. Accessed 1st November, 2020.
16. Surgeons RACo. Professional Development Program for Supervisors. https://www.surgeons.org/Education/professional-development/supervisors. Accessed.
17. Yaqoob MM, Ashman N. Kidney and urinary tract disease. In: Kumar P, Clark M, eds. *Clinical Medicine.* 9 ed. London: Elsevier ltd.; 2017.
18. Sadler DR. Formative assessment and the design of instructional systems. *Instructional Science.* 1989;18(2):119-144.
19. Boud D, Molloy E. What is the problem with feedback? In: Boud D, Molloy E, eds. *Feedback in higher and professional education.* London: Routledge; 2013:1-10.

20. Kaufman DM, Mann KV. Teaching and learning in medical education: How theory can inform practice. In: Swanwick T, ed. *Understanding medical education. Evidence, theory and practice.* 2nd ed. Oxford: Wiley Blackwell; 2014:7-29.

21. Higgs J. Ways of knowing for clinical practice. In: Delany C, Molloy E, eds. *Clinical education in the health professions.* Sydney: Elsevier; 2009.

22. Morris C, Blaney D. Work-based learning In: Swanwick T, ed. *Understanding medical education: evidence, theory and practice.* 2nd ed. Oxford: The Association for the study of Medical Education 2014:97-109.

23. Molloy E, Van de Ridder M. Reworking feedback to to build better work. In: Delany C, Molloy E, eds. *Learning and teaching in clinical contexts.* Sydney: Elsevier; 2018:305-320.

24. Vickery AW, Lake FR. Teaching on the run tips 10: giving feedback. *Med J Aust.* 2005;183(5):267-268.

25. Jolly B, Boud D. Written feedback. In: Boud D, Molloy E, eds. *Feedback in higher and professional education.* London: Routledge; 2013.

26. Carless D, Boud D. The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education.* 2018;43(8):1315-1325.

27. Carless D, Winstone N. Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education.* 2020:1-14.

28. Noble C, Billett S, Armit L, et al. "It's yours to take": generating learner feedback literacy in the workplace. *Advances in health sciences education : theory and practice.* 2019.

29. Carless D. Trust and its role in facilitating dialogic feedback. In: Boud D, Molloy E, eds. *Feedback in higher and professional education.* London: Routledge; 2013:90-103.

30. Molloy E, Borrell-Carrio F, Epstein R. The impact of emotions in feedback. In: Boud D, Molloy E, eds. *Feedback in higher and professional education. .* London: Routledge; 2013:50-71.

31. Bynum WE, Haque TM. Risky Business: Psychological Safety and the Risks of Learning Medicine. *Journal of Graduate Medical Education.* 2016;8(5):780-782.

32. Ten Cate OT. Why receiving feedback collides with self determination. *Advances in Health Sciences Education.* 2013;18(4):845-849.

33. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev.* 2012(6):CD000259.

34. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher.* 2006;28(2):117-128.

35. 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(3):275-288.

36. Hewson MG, Little ML. Giving Feedback in Medical Education: Verification of Recommended Techniques. *Journal of General Internal Medicine.* 1998;13(2):111-116.

37. Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Medical Education.* 2005;39(5):497-504.

38. 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(2):201-215.

39. Mattick K, Brennan N, Briscoe S, Papoutsi C, Pearson M. Optimising feedback for early career professionals: a scoping review and new framework. *Med Educ.* 2019;53(4):355-368.

40. Creswell JW. Mixed methods design. In: *Educational research: planning, conducting, and evaluating quantitative and qualitative research.* 4 ed. Boston, MA, : Pearson Education Inc.; 2012:534-575.

41. Bazeley P. *Qualitative data analysis : practical strategies / Pat Bazeley.* London ; Thousand Oaks, Calif: SAGE; 2013.

42. Creswell JW. Specifying a purpose and research questions or hypotheses. In: *Educational research: planning, conducting, and evaluating quantitative and qualitative research.* 4 ed. Boston, MA, USA: Pearson Education Inc.; 2012:109-139.

43. Vygotsky LS. Interation between learning and development. In: Cole M, John-Steiner V, Scribner S, Souberman E, eds. *Mind and Society.* Cambridge Massachusetts: Harvard University Press; 1978:77-91.

44. Trent SC, Artiles AJ, Englert CS. From Deficit Thinking to Social Constructivism: A Review of Theory, Research, and Practice in Special Education. *Review of Research in Education.* 1998;23:277-307.

45. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126(5):376-380.

46. Carraccio C, Englander R, Van Melle E, et al. Advancing Competency-Based Medical Education: A Charter for Clinician-Educators. *Academic Medicine.* 2016;91(5):645-649.

47. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Medical Teacher.* 2012;34(3):205-214.

48. Holmboe ES. Realizing the promise of competency-based medical education. *Academic Medicine.* 2015;90(4):411-413.

49. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Journal of the American Medical Association.* 2009;302(12):1316-1326.

50. Archer JC. State of the science in health professional education: effective feedback. *Medical Education.* 2010;44:101-108.

51. Bing-You RG, Hayes V, Varaklis K, Trowbridge R, Kemp H, McKelvy D. Feedback for learners in medical education: what is known? A scoping review. *Academic Medicine.* 2017;92(9):1346-1354.

52. Johnson CE, Keating JL, Boud DJ, et al. Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement. *BMC Medical Education.* 2016;16(1):96.

53. Watling C, Ginsburg S. Assessment, feedback and the alchemy of learning. *Medical Education.* 2019;53(1):76-85.

54. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *The American psychologist.* 2002;57(9):705-717.

55. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Academic Medicine.* 2015;90(5):609-614.

56. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Medical Teacher.* 2011;33(12):961-973.

57. Johnson CE, Keating JL, Farlie MK, Kent F, Leech M, Molloy EK. Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. *BMC Medical Education.* 2019;19(1):129.

58. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med.* 2009;151(4):264-269, w264.

59. Higgins JPT, Churchill R, Chandler J, Cumpston MSe. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). In: Available from www.training.cochrane.org/handbook: Cochrane; 2017.

60. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *British Medical Journal.* 2019;366:l4898.

61. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List. *Journal of Clinical Epidemiology.* 1998;51(12):1235-1241.

62. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal.* 2011;343.

63. Sterne JAC, Egger M, Moher D, Boutron Ie. Chapter 10: Addressing reporting biases. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017)*. Available from www.training.cochrane.org/handbook: Cochrane; 2017.

64. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *British Medical Journal.* 2011;342:d549.

65. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] The Cochrane Collaboration. www.handbook.cochrane.org. Published 2011. Accessed.

66. Schünemann HJ, Oxman AD, Vist GE, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017)*. Available from www.training.cochrane.org/handbook: Cochrane; 2017.

67. GRADE guidelines Journal of Clinical Epidemiology series https://www.gradeworkinggroup.org/. Accessed2020.

68. Ahlborg L, Weurlander M, Hedman L, et al. Individualized feedback during simulated laparoscopic training:a mixed methods study. *International Journal of Medical Education.* 2015;6:93-100.

69. Boyle E, O'Keeffe DA, Naughton PA, Hill AD, McDonnell CO, Moneley D. The importance of expert feedback during endovascular simulator training. *Journal of Vascular Surgery.* 2011;54(1):240-248.e241.

70. Olms C, Jakstat HA, Haak R. The Implementation of Elaborative Feedback for Qualitative Improvement of Shade Matching-A Randomized Study. *Journal of Esthetic & Restorative Dentistry.* 2016;28(5):277-286.

71. Pavo N, Goliasch G, Nierscher FJ, et al. Short structured feedback training is equivalent to a mechanical feedback device in two-rescuer BLS: a randomised simulation study. *Scandinavian Journal of Trauma, Resuscitation & Emergency Medicine.* 2016;24:70.

72. Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP. Comprehensive Surgical Coaching Enhances Surgical Skill in the Operating Room: A Randomized Controlled Trial. *Annals of Surgery.* 2015;262(2):205-212.

73. Kroft J, Ordon M, Po L, et al. Preoperative Practice Paired With Instructor Feedback May Not Improve Obstetrics-Gynecology Residents' Operative Performance. *Journal of Graduate Medical Education.* 2017;9(2):190-194.

74. Soucisse ML, Boulva K, Sideris L, Drolet P, Morin M, Dube P. Video Coaching as an Efficient Teaching Method for Surgical Residents-A Randomized Controlled Trial. *Journal of Surgical Education.* 2017;74(2):365-371.

75. Xeroulis GJ, Park J, Moulton CA, Reznick RK, Leblanc V, Dubrowski A. Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery.* 2007;141(4):442-449.

76. O'Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surgical Endoscopy.* 2008;22(7):1614-1619.

77. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *American Journal of Medicine.* 1983;75(3):465-470.

78. Vafaei A, Heidari K, Hosseini MA, Alavi-Moghaddam M. Role of feedback during evaluation in improving emergency medicine residents' skills; an experimental study. *Emergency.* 2017;5 (1) (no pagination)(e28).

79. Lake FR, Hamdorf JM. Teaching on the run tips 5: teaching a skill. *Medical Journal of Australia.* 2004;181(6):327-328.

80. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *Journal of Clinical Epidemiology.* 2013;66(2):151-157.

81. Al-Jundi W, Elsharif M, Anderson M, Chan P, Beard J, Nawaz S. A Randomized Controlled Trial to Compare e-Feedback Versus "Standard" Face-to-Face Verbal Feedback to Improve the Acquisition of Procedural Skill. *Journal of Surgical Education.* 2017;74(3):390-397.

82. Baldwin LJ, Jones CM, Hulme J, Owen A. Use of the learning conversation improves instructor confidence in life support training: An open randomised controlled cross-over trial comparing teaching feedback mechanisms. *Resuscitation.* 2015;96:199-207.

83. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. *Medical Education.* 2006;40(8):746-749.

84. Bosse HM, Mohr J, Buss B, et al. The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. *BMC Medical Education.* 2015;15:22.

85. Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine.* 2007;161(1):44-49.

86. DeLucenay AJ, Conn KM, Corigliano A. An evaluation of the impact of immediate compared to delayed feedback on the development of counselling skills in pharmacy students. *Pharmacy Education.* 2017;17(1):322-328.

87. Lee JY, McDougall EM, Lineberry M, Tekian A. Optimizing the Timing of Expert Feedback During Simulation-Based Spaced Practice of Endourologic Skills. *Simulation in Healthcare: The Journal of The Society for Medical Simulation.* 2016;11(4):257-263.

88. Ozcakar N, Mevsim V, Guldal D, et al. Is the use of videotape recording superior to verbal feedback alone in the teaching of clinical skills? *BMC Public Health.* 2009;9:474.

89. Strandbygaard J, Bjerrum F, Maagaard M, et al. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial. *Annals of Surgery.* 2013;257(5):839-844.

90. Ridder J, Peters C, Stokking K, Ru J, Cate O. Framing of feedback impacts student's satisfaction, self-efficacy and performance. *Advances in Health Sciences Education.* 2015;20(3):803-816.

91. van de Ridder J, Berk FC, Stokking KM, ten Cate OT. Feedback providers' credibility impacts students' satisfaction with feedback and delayed performance. *Medical Teacher.* 2015;37(8):767-774.

92. Rogers DA, Boehler ML, Schwind CJ, Meier AH, Wall JC, Brenner MJ. Engaging medical students in the feedback process. *American Journal of Surgery.* 2012;203(1):21-25.

93. Manzone J, Tremblay L, You-Ten KE, Desai D, Brydges R. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education.* 2014;48(4):430-440.

94. Sox CM, Dell M, Phillipi CA, Cabral HJ, Vargas G, Lewin LO. Feedback on oral presentations during pediatric clerkships: a randomized controlled trial. *Pediatrics.* 2014;134(5):965-971.

95. Backstein D, Agnidis Z, Sahdu R, MacRae H. Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery.* 2005;48(3):195-200.

96. Rudolph JW, Simon R, Dufresne RL, Raemer DB. There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. *Simulation in Healthcare: The Journal of The Society for Medical Simulation.* 2016;1(1):49-55.

97. Cohen J. *Statistical power analysis for the behavioral sciences.* Hillsdale, N.J.: L. Erlbaum Associates; 1988.

98. Shute VJ. Focus on Formative Feedback. *Review of Educational Research.* 2008;78(1):153-189.

99. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Medical Education.* 2016;50(9):933-942.

100. Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Medical Education.* 2007;41(6):583-591.

101. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education.* 2012;17(1):15-26.

102. Fitts PM, Posner MI. *Human Performance.* Belmont CA: Brooks/Cole; 1967.

103. Li Q, Ma EL, Liu J, Fang LQ, Xia T. Pre-training evaluation and feedback improve medical students' skills in basic life support. *Medical Teacher.* 2011;33(10):e549-555.

104. Elliott SN. *Educational psychology: effective teaching, effective learning.* 3rd ed. Boston: McGraw-Hill; 2000.

105. Wadsworth BJ. *Piaget's theory of cognitive and affective development: Foundations of constructivism.* 5th ed. White Plains: Longman Publishing; 1996.

106. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educational Psychology Review.* 1998;10(3):251-296.

107. Deci EL, Ryan RM. The 'What' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry.* 2000;11:227-268.

108. Belfield C, Thomas H, Bullock A, Eynon R, Wall D. Measuring effectiveness for best evidence medical education: a discussion. *Med Teach.* 2001;23(2):164-170.

109. Powell C. The Delphi technique: myths and realities. *Journal of Advanced Nursing.* 2003;41.

110. Bond S, Bond J. A Delphi study of clinical nursing research priorities. *J Adv Nurs.* 1982;7.

111. Nicol D. Resituating feedback from the reactive to the proactive. In: Boud D, Molloy E, eds. *Feedback in higher and professional education.* Oxford: Routledge; 2013:34-49.

112. Kilminster S, Cottrell D, Grant J, Jolly B. AMEE Guide No. 27: effective educational and clinical supervision. *Med Teach.* 2007;29.

113. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education.* 2006;31(2):199-218.

114. Blatt B, Confessore S, Kallenberg G, Greenberg L. Verbal interaction analysis: viewing feedback through a different lens. *Teaching and Learning in Medicine.* 2008;20(4):329-333.

115. Molloy E. Time to pause: feedback in clinical education. In: Delany C, Molloy E, eds. *Clinical education in the health professions.* Sydney: Elsevier; 2009:128-146.

116. Salerno SM, O'Malley PG, Pangaro LN, Wheeler GA, Moores LK, Jackson JL. Faculty development seminars based on the one-minute preceptor improve feedback in the ambulatory setting. *Journal of General Internal Medicine.* 2002;17(10):779-787.

117. Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *Journal of General Internal Medicine.* 2004;19(5 Pt2):558-561.

118. Frye AW, Hollingsworth MA, Wymer A, Hinds MA. Dimensions of feedback in clinical teaching: a descriptive study. *Academic Medicine.* 1996;71(1):S79-81.

119. Pelgrim EAM, Kramer AWM, Mokkink HGA, Vleuten CPM. Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: an observational study. *BMC Medical Education.* 2012;12(12):97.

120. Hamburger EK, Cuzzi S, Coddington DA, et al. Observation of resident clinical skills: outcomes of a program of direct observation in the continuity clinic setting. *Academic Pediatrics.* 2011;11(5):394–402.

121. Huang WY, Dains JE, Monteiro FM, Rogers JC. Observations on the teaching and learning occurring in offices of community-based family and community medicine clerkship preceptors. *Family Medicine.* 2004;36(2):131-136.

122. Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM. Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. *Acad Med.* 1998;73.

123. Moss HA, Derman PB, Clement RC. Medical student perspective: working toward specific and actionable clinical clerkship feedback. *Medical Teacher.* 2012;34.

124. Molloy E, Boud D. Feedback models for learning, teaching and performance. In: Spector JM, Merrill MD, Elen J, Bishop MJ, eds. *Handbook of research on*

educational communications and technology. New York: Springer Science + Business Media; 2014.

125. Pangaro L, Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach.* 2013;35.

126. Booth A. Unpacking your literature search toolbox: on search styles and tactics. *Health Info Libr J.* 2008;25.

127. Miles M, Huberman AJS. *Qualitative data analysis: a methods sourcebook.* Los Angeles: Sage; 2014.

128. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use.* Oxford University Press: New York; 2003.

129. Duffield C. The Delphi technique: a comparison of results obtained using two expert panels. *International Journal of Nursing Studies.* 1993;30.

130. Riddle DL, Stratfrod PW, Singh JA, Strand CV. Variation in outcome measures in hip and knee arthroplasty clinical trials: a proposed approach to achieving consensus. *The Journal of Rheumatology.* 2009;36.

131. van de Ridder JM, Stokking KM, McGaghie WC, ten Cate OT. What is feedback in clinical education? *Medical Education.* 2008;42(2):189-197.

132. Boud D, Molloy E. *Feedback in Higher and Professional Education: Understanding it and doing it well.* NY USA: Routledge; 2013.

133. Harvard Business Review Guide to Giving Effective Feedback. Boston MA: Harvard Buiness Review Press; 2012. In.

134. Swanwick T. *Understanding medical education: evidence, theory and practice.* 1st ed. Oxford, UK: Association for the study of medical education; 2010.

135. Cantillon P, Hutchinson L, Wood D. ABC of learning and teaching in medicine. In. 2nd ed. Oxford UK: Blackwell Publishing Ltd; 2010.

136. Cooper N, Forrest K. *Essential guide to educational supervision in postgraduate medical education.* Oxford: Wiley; 2009.

137. Wind LA, Dalen J, Muijtjens AM, Rethans JJ. Assessing simulated patients in an educational setting: the MaSP (Maastricht assessment of simulated patients). *Medical Education.* 2004;38.

138. Perera J, Perera J, Abdullah J, Lee N. Training simulated patients: evaluation of a training approach using self-assessment and peer/tutor feedback to improve performance. *BMC Medical Education.* 2009;9.

139. May W, Fisher D, Souder D. Development of an instrument to measure the quality of standardized/simulated patient verbal feedback. *Medical Education Development.* 2012;2.

140. Schlegel C, Woermann U, Rethans JJ, Vleuten C. Validity evidence and reliability of a simulated patient feedback instrument. *BMC Medical Education.* 2012;12.

141. Bouter S, Weel-Baumgarten EMDP, Bolhuis SP. Construction and validation of the nijmegen evaluation of the simulated patient (nesp): assessing simulated patients' ability to role-play and provide feedback to students. *Acad Med.* 2013;88.

142. Gururaja RP, Yang T, Paige JT, Chauvin SW. Examining the effectiveness of debriefing at the point of care in simulation-based operating room team training. In: Henriksen K, Battles JB, Keyes MA, Gary ML, eds. *Advances in patient safety: new directions and alternative approaches (vol 3: performance and tools).* Rockville: Agency for Healthcare Research and Quality; 2008.

143. Brett-Fleegler M, Rudolph J, Eppich W, et al. Debriefing assessment for simulation in healthcare: development and psychometric properties. *Simulation in Healthcare.* 2012;7(5):288-294.

144. Arora SP, Ahmed MMPH, Paige JMD, et al. Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. . *Annals of Surgery.*256(6):982-988.

145. Reiter HI, Rosenfeld J, Nandagopal K, Eva KW. Do clinical clerks provide candidates with adequate formative assessment during objective structured clinical examinations? *Adv in Health Sci Educ.* 2004;9.

146. Moineau G, Power B, Pion A-MJ, Wood TJ, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. *Med Educ.* 2011;45.

147. Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Peters WG, Coppenolle L, Scherpbier AJJA. Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Adv in Health Sci Educ.* 2010;15.

148. Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity. *Med Teach.* 2008;30.

149. C. F, Bolhuis S, R. G, et al. Evaluation and feedback for effective clinical teaching in postgraduate medical education: Validation of an assessment instrument incorporating the CanMEDS roles. *Medical Teacher.* 2012;34(11):893–901.

150. Kaufman DM. Applying educational theory in practice. In: Cantillon P, Wood D, eds. *ABC of learning and teaching in medicine.* Oxford: Blackwell Publishing Ltd; 2010.

151. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci.* 1988;12.

152. Butler DL, Winne PH. Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research.* 1995;65(3):245-281.

153. Rudolph JW, Simon R, Rivard P, Dufresne RL, Raemer DB. Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Anesthesiology Clinics.* 2007;25.

154. Silverman J, Kurtz S. The Calgary-Cambridge approach to communication skills teaching ii: the set-go method of descriptive feedback. *Education for General Practice.* 1997;8(7):288-299.

155. Brett JF, Atwater LE. 360degree feedback: accuracy, reactions, and perceptions of usefulness. *J Appl Psychol.* 2001;86.

156. Kernis MH, Johnson EK. Current and typical self-appraisals: differential responsiveness to evaluative feedback and implications for emotions. *J Res Pers.* 1990;24.

157. Dweck CS. Motivational processes affecting learning. *American Psychologist.* 1986;41.

158. Sargeant J, McNaughton E, Mercer S, Murphy D, Sullivan P, Bruce DA. Providing feedback: exploring a model (emotion, content, outcomes) for facilitating multisource feedback. *Medical Teacher.* 2011;33(9):744-749.

159. Rudolph JW, Simon R, Raemer DB, Eppich WJ. Debriefing as formative assessment: closing performance gaps in medical education. *Academic Emergency Medicine.* 2008;15(11):1010-1016.

160. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach.* 2010;32.

161. Miller GE. The assessment of clinical skills/competence/performance. *Academic Medicine.* 1990;65(9 Suppl):S63-67.

162. Ericsson KA. Acquisition and maintenance of medical expertise: a perspective from the expert-performance approach with deliberate practice. *Academic Medicine.* 2015;90(11):1471-1486.

163. Bardella IJ, Janosky J, Elnicki DM, Ploof D, Kolarik R. Observed versus reported precepting skills: teaching behaviours in a community ambulatory clerkship. *Medical Education.* 2005;39(10):1036-1044.

164. Hekelman FP, Vanek E, Kelly K, Alemagno S. Characteristics of family physicians' clinical teaching behaviors in the ambulatory setting: A descriptive study. *Teaching and Learning in Medicine.* 1993;5(1):18-23.

165. Urquhart LM, Ker JS, Rees CE. Exploring the influence of context on feedback at medical school: a video-ethnography study. *Advances in health sciences education : theory and practice.* 2018;23(1):159-186.

166. Ramani S, Konings KD, Ginsburg S, van der Vleuten CPM. Twelve tips to promote a feedback culture with a growth mind-set: Swinging the feedback pendulum from recipes to relationships. *Medical Teacher.* 2018:1-7.

167. 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(12):1698-1706.

168. Sargeant J, Lockyer JM, Mann K, et al. The R2C2 Model in Residency Education: How Does It Foster Coaching and Promote Feedback Use? *Academic Medicine.* 2018;93(7):1055-1063.

169. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education.* 2010;15(5):625-632.

170. Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Medical Teacher.* 2011;33(11):919-927.

171. Fernando N, Cleland J, McKenzie H, Cassar K. Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments. *Medical Education.* 2008;42(1):89-95.

172. Johnson CE, Molloy EK. Building evaluative judgement through the process of feedback. In: Boud D, Ajjawi R, Dawson P, Tai J, eds. *Developing evaluative judgement in higher education. Assessment for knowing and producing quality work.* London: Routledge; 2018:166-175.

173. Grenny J, Patterson K, Maxfield D, McMillan R, Switzler A. *Infuencer The Science of Leading Change.* 2nd ed. New York: MC Graw Hill Education; 2013.

174. Patterson K, Grenny J, McMillan R, Switzler A. *Crucial conversations: Tools for talking when the stakes are high.* 2nd ed. New York: McGraw-Hill 2012.

175. Voyer S, Cuncic C, Butler DL, MacNeil K, Watling C, Hatala R. Investigating conditions for meaningful feedback in the context of an evidence-based feedback programme. *Medical Education.* 2016;50(9):943-954.

176. LaDonna KA, Hatala R, Lingard L, Voyer S, Watling C. Staging a performance: learners' perceptions about direct observation during residency. *Medical Education.* 2017;51(5):498-510.

177. Flinn JT, Miller A, Pyatka N, Brewer J, Schneider T, Cao CG. The effect of stress on learning in surgical skill acquisition. *Medical Teacher.* 2016;38(9):897-903.

178. Irby DM. Excellence in clinical teaching: knowledge transformation and development required. *Med Educ.* 2014;48(8):776-784.

179. Knight J. *Focus on teaching: Using video for high-impact instruction.* Thousand Oaks, CA.: Corwin; 2014.

180. Tai JH, Canny BJ, Haines TP, Molloy EK. The role of peer-assisted learning in building evaluative judgement: opportunities in clinical medical education. *Advances in Health Sciences Education.* 2016;21(3):659-676.

181. Winstone NE, Nash RA, Parker M, Rowntree J. Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes *Educational Psychologist.* 2017;52(1):17-37.

182. Delany C, Molloy E. Critical reflection in clinical education: beyond the 'swampy lowlands'. In: Delany C, Molloy E, eds. *Clinical Education in the Health Professions.* 1st ed. Sydney, Australia: Elsevier; 2009:3-24.

183. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999;77(6):1121-1134.

184. Rees CE, Bullock A, Mattick KL, Monrouxe LV. Using workplace-learning narratives to explore evaluative judgment in action. In: Boud D, Ajjawi R, Dawson P, Tai J, eds. *Developing evaluative judgement in higher education. Assessment for knowing and producing quality work.* London: Routledge; 2018:176-185.

185. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine.* 2005;80(10 Suppl):S46-54.

186. Bearman M, Molloy E, Ajjawi R, Keating J. 'Is there a Plan B?': clinical educators supporting underperforming students in practice settings. *Teaching in Higher Education.* 2013;18(5):531-544.

187. Clouder L. 'Being responsible': students' perspectives on trust, risk and work-based learning. *Teaching in Higher Education.* 2009;14(3):289-301.

188. Lefevre M. Facilitating practice learning and assessment: The influence of relationship. *Social Work Education.* 2005;24(5):565-583.

189. Tai J, Sevenhuysen S. The role of peers in developing evaluative judgement. In: Boud D, Ajjawi R, Dawson P, Tai J, eds. *Developing evaluative judgement in higher education. Assessment for knowing and producing quality work.* London: Routledge; 2018:156-165.

190. Carless D. Differing perceptions in the feedback process. *Studies in Higher Education.* 2006;31(2):219-233.

191. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Academic Medicine.* 2005;80(10):S84-S87.

192. Edmondson AC. Learning from Mistakes is Easier Said Than Done: Group and Organizational Influences on the Detection and Correction of Human Error. *The Journal of Applied Behavioral Science.* 1996;32(1):5-28.

193. Edmondson AC. Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly.* 1999;44(2):350-383.

194. Rosenbaum L. Cursed by Knowledge — Building a Culture of Psychological Safety. *New England Journal of Medicine.* 2019;380(8):786-790.

195. Bynum WE, Artin AR. Why we should strive for emotional candour in medical education, too. *Medical Education.* 2019;53(7):745-746.

196. Rudolph JW, Raemer DB, Simon R. Establishing a Safe Container for Learning in Simulation: The Role of the Presimulation Briefing. *Simulation in Healthcare.* 2014;9(6):339-349.

197. Kolbe M, Eppich W, Rudolph J, et al. Managing psychological safety in debriefings: a dynamic balancing act. *BMJ Simulation and Technology Enhanced Learning.* 2019:bmjstel-2019-000470.

198. Silverman J, Kurtz S, Draper J. *Skills for communicating with patients.* 3rd ed. London: Radcliffe Publishing; 2013.

199. Watling CJ. Unfulfilled promise, untapped potential: feedback at the crossroads. *Medical Teacher.* 2014;36(8):692-697.

200. Wood D. Formative assessment. In: Swanwick T, ed. *Understanding medical education. Evidence, theory and practice.* 1st ed. Oxford: Wiley-Blackwell; 2010.

201. van der Leeuw RM, Teunissen PW, van der Vleuten CPM. Broadening the Scope of Feedback to Promote Its Relevance to Workplace Learning. *Academic Medicine.* 2018;93(4):556-559.

202. Johnson CE, Keating JL, Leech M, et al. The Feedback Quality Instrument: A guide for health professional educators in fostering learner-centred discussions. *Under submission* 2020.

203. Dawson P, Ajjawi R, Boud D, Tai J. Introduction: what is evaluative judgement? In: Boud D, Ajjawi R, Dawson P, Tai J, eds. *Developing evaluative judgement in higher education. Assessment for knowing and producing quality work.* London: Routledge; 2018:1-4.

204. Tai JH, Ajjawi R, Boud D, Dawson P, Panadero E. Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education.* 2017: https://doi.org/10.1007/s10734-10017-10220-10733.

205. Ajjawi R, Boud D. Researching feedback dialogue: an interactional analysis approach. *Assessment & Evaluation in Higher Education.* 2017;42(2):252-265.

206. Armson H, Lockyer JM, Zetkulic M, Könings KD, Sargeant J. Identifying coaching skills to improve feedback use in postgraduate medical education. *Medical Education.* 2019;53(5):477-493.

207. Molloy E, Bearman M. Embracing the tension between vulnerability and credibility: 'intellectual candour' in health professions education. *Med Educ.* 2019;53(1):32-41.

208. Watling CJ, Lingard L. Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Advances in health sciences education : theory and practice.* 2012;17(2):183-194.

209. Winstone NE, Nash RA, Rowntree J, Parker M. 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. *Studies in Higher Education.* 2017;42(11):2026-2041.

210. Sargeant J, Armson H, Chesluk B, et al. The processes and dimensions of informed self-assessment: a conceptual model. *Academic Medicine.* 2010;85(7):1212-1220.

211. Tschannen-Moran M. *Trust matters: Leadership for successful schools.* San Francisco: Jossey Bass; 2004.

212. Riskin A, Erez A, Foulk TA, et al. The Impact of Rudeness on Medical Team Performance: A Randomized Trial. *Pediatrics.* 2015;136(3):487-495.

213. Katz D, Blasius K, Isaak R, et al. Exposure to incivility hinders clinical performance in a simulated operative crisis. *BMJ Qual Saf.* 2019;28(9):750-757.

214. Krogh K, Bearman M, Nestel D. “Thinking on your feet”—a qualitative study of debriefing practice. *Advances in Simulation.* 2016;1(1):12.

215. Silverman J, Kurtz S, Draper J. Building the relationship. In: *Skills for communicating with patients.* 3rd ed. London: Radcliffe Publishing 2013:118-148.

216. Harrison CJ, Konings 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 : theory and practice.* 2015;20(1):229-245.

217. Castanelli DJ, Jowsey T, Chen Y, Weller JM. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anesthesia training. *Can J Anaesth.* 2016;63(12):1345-1356.

218. Panhwar MS, Kalra A. Breaking Down the Hierarchy of Medicine: The airline industry has taken the lead to improve communications for pilots, it is now time for medicine to follow with physicians. *European Heart Journal.* 2019;40(19):1482-1483.

219. Crowe S, Clarke N, Brugha R. ‘You do not cross them’: Hierarchy and emotion in doctors' narratives of power relations in specialist training. *Social Science & Medicine.* 2017;186:70-77.

220. Palincsar AS. Social constructivist perspectives on teaching and learning. *Annual Review of Psychology.* 1998;49(1):345-375.

221. Boeije H. A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity.* 2002;36(4):391-409.

222. Haidet P, Paterniti DA. "Building" a History Rather Than "Taking" One: A Perspective on Information Sharing During the Medical Interview. *Archives of Internal Medicine.* 2003;163(10):1134-1140.

223. Platt FW, Platt CM. Two Collaborating Artists Produce a Work of Art: The Medical Interview. *Archives of Internal Medicine.* 2003;163(10):1131-1132.

224. Zimmermann C, Del Piccolo L, Finset A. Cues and concerns by patients in medical consultations: A literature review. *Psychological Bulletin.* 2007;133(3):438-463.

225. Urrieta L. Figured Worlds and Education: An Introduction to the Special Issue. *The Urban Review.* 2007;39(2):107-116.

226. Silverman J, Kurtz S, Draper J. Defining what to teach and learn. In: *Skills for communicating with patients.* 3rd ed. London: Radcliffe Publishing 2013:7-34.

227. Silverman J, Kurtz S, Draper J. Initiating the session. In: *Skills for communicating with patients.* 3rd ed. London Radcliffe Pulishing; 2013:35-58.

228. Bearman M, Molloy E. Intellectual streaking: The value of teachers exposing minds (and hearts). *Medical Teacher.* 2017;39(12):1284-1285.

229. Billett S. Errors and learing from errors at work. In: Bauer J, Hartteis C, eds. *Human Falibility: The Ambiguity of Errors for Work and Learning.* Dordrecht: Springer; 2012:17-32

230. Ramani S, Post SE, Könings K, Mann K, Katz JT, van der Vleuten C. “It's Just Not the Culture”: A Qualitative Study Exploring Residents' Perceptions of the Impact of Institutional Culture on Feedback. *Teaching and Learning in Medicine.* 2017;29(2):153-161.

231. Iedema R, Carroll K, Collier A, Su-Yin H, Mesman J, Wyer M. *Video-reflexive ethnography in health research and healthcare improvement.* 1st ed. Boca Raton, FL: CRC Press; 2019.

232. Linacre JM. *Many-faceted Rasch measurement.* 2nd ed. Chicago: MESA Press; 1994.

233. Boone WJ, Staver JR, Yale MS. *Rasch Analysis in the Human Sciences.* Dordrecht: Springer Netherlands; 2014.

234. Bond TG, Fox CM. *Applying the Rasch model: fundamental measurement in the human sciences.* 3rd ed. New York, NY: Routledge; 2015.

235. Johnson CE, Weerasuria MP, Keating JL. Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis. *BMJ Open.* 2020;10(3):e030672.

236. Pelgrim EA, Kramer AW, Mokkink HG, van der Vleuten CP. The process of feedback in workplace-based assessment: organisation, delivery, continuity. *Medical Education.* 2012;46(6):604-612.

237. Carless D. Double duty, shared responsibilities and feedback literacy. *Perspectives on Medical Education.* 2020.

238. Molloy E, Ajjawi R, Bearman M, Noble C, Rudland J, Ryan A. Challenging feedback myths: Values, learner involvement and promoting effects beyond the immediate task. *Medical Education.* 2019;0(0):1-7.

239. Nash RA, Winstone NE. Responsibility-Sharing in the Giving and Receiving of Assessment Feedback. *Frontiers in Psychology.* 2017;8(1519).

240. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health.* 2018;6(149).

241. Pett MA, Lackey NR, Sullivan JJ. Designing and testing the instrument. In: *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Thousand Oaks, California: SAGE Publications; 2003:29-49.

242. Johnson CE, Keating JL, Molloy EK. Psychological safety in feedback: What does it look like and how can educators work with learners to foster it? *Medical Education.* 2020;54(6):559-570.

243. Pallant JF. Factor analysis. In: *SPSS survival manual.* Sydney: Allen & Unwin; 2016:182-203.

244. Pett MA, Lackey NR, Sullivan JJ. Making Sense of Factor Analysis. In: Thousand Oaks, California: SAGE publications; 2003.

245. Yong AG, Pearce S. A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology.* 2013;9(2):79-94.

246. Tavakol M, Dennick R. Psychometric evaluation of a knowledge based examination using Rasch analysis: An illustrative guide: AMEE Guide No. 72. *Medical Teacher.* 2013;35(1):e838-e848.

247. Wolcott MD, Zeeman JM, Cox WC, McLaughlin JE. Using the multiple mini interview as an assessment strategy within the first year of a health professions curriculum. *BMC Medical Education.* 2018;18(1):92.

248. Tabachnick BG, Fidell LS. *Using multivariate statistics.* 6th ed. Boston: Pearson Education; 2013.

249. Molloy E, Boud D, Henderson M. Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education.* 2020;45(4):527-540.

250. Dweck CS, Yeager DS. Mindsets: A View From Two Eras. *Perspectives on Psychological Science.* 2019;14(3):481-496.

251. Cook DA, Artino AR, Jr. Motivation to learn: an overview of contemporary theories. *Medical Education.* 2016;50(10):997-1014.

252. Molloy E, Boud D. Seeking a different angle on feedback in clinical education: the learner as seeker, judge and user of performance information. *Medical Education.* 2013;47(3):227-229.

253. Mpotos N, Yde L, Calle P, et al. Retraining basic life support skills using video, voice feedback or both: a randomised controlled trial. *Resuscitation.* 2013;84(1):72-77.

254. Clement T, Howard D, Lyon E, Silverman J, Molloy E. Video-triggered professional learning for general practice trainers: using the 'cauldron of practice' to explore teaching and learning. *Education for Primary Care.* 2020;31(2):112-118.

255. Rooney D, Boud D. Toward a Pedagogy for Professional Noticing: Learning through Observation. *Vocations and Learning.* 2019;12(3):441-457.

256. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. *Qualitative Health Research.* 1993;3(4):430-441.

257. Rasch G. *Probabilistic models for some intelligence and attainments tests.* Chicago: University of Chicago Press 1960.

258. Boone WJ. Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sci Educ.* 2016;15(4):rm4.

259. Granger C. Rasch analysis is important to understand and use for measurement. *Rasch Measurement Transactions.* 2008;21(3):1122-1123.

260. Bond T, Fox CM. Measuring facets beyond ability and difficulty. In: *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* 3rd ed. London, United Kingdom: Taylor & Francis Group; 2015:167-187.

261. Boone WJ, Staver JR, Yale MS. Multifaceted rasch measurement. In: *Rasch Analysis in the Human Sciences.* Dordrecht: Springer Netherlands; 2014:423-448.

262. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions.* 1994;8(3):370.

263. Bond TG, Fox CM. Rasch Model Requirements: Model Fit and Unidimensionality. In: *Applying the Rasch Model : Fundamental Measurement in the Human Sciences.* New York, NY: Taylor & Francis Group; 2015.

264. Linacre J. Reliability, Separation and Strata: Percentage of Sample in Each Level. *Rasch Measurement Transactions.* 2013;26(4):1399.

265. Bartlett MS. A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society* 1954;16:296-298.

266. Kaiser HF. An index of factorial simplicity. *Psychometrika.* 1974;39(1):31-36.

267. Hutcheson GD, Sofroniou N. *The Multivariate Social Scientist: an Introduction to Generalized Linear Models.* London: Sage Publications; 1999.

268. Catell RB. The scree test for number of factors. *Multivariate Behavioral Research* 1966;1:245-276.

269. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika.* 1965;30:179-185.

270. Yen PH, Leasure AR. Use and Effectiveness of the Teach-Back Method in Patient Education and Health Outcomes. *Federal Practitioner.* 2019;36(6):284-289.

Appendices

1. Johnson CE, Weerasuria MP, Keating JL. Effect of face-to-face verbal feedback compared with no or alternative feedback on the objective workplace task performance of health professionals: a systematic review and meta-analysis. BMJ Open. 2020;10(3):e030672 (Chapter 2)
2. Johnson CE, Keating JL, Boud DJ, Dalton M, Kiegaldie D, Hay M, McGrath B, McKenzie WA, Nair KBR, Nestel D, Palermo C, Molloy EK. Identifying educator behaviours for high quality verbal feedback in health professions education: literature review and expert refinement. BMC Medical Education. 2016;16(1):96. (Chapter 3)
3. Johnson CE, Keating JL, Farlie MK, Kent F, Leech M, Molloy EK. Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. BMC Medical Education. 2019;19(1):129. (Chapter 4)
4. Johnson CE, Keating JL, Molloy EK. Psychological safety in feedback: What does it look like and how can educators work with learners to foster it? Medical Education. 2020;54(6):559-570. (Chapter 6)
5. Johnson CE, Watling CJ, Keating JL, Molloy EK. Effective feedback conversations in clinical practice. In: Nestel D, Reedy G, McKenna L, Gough S, eds. Clinical Education for the Health Professions: Theory and Practice. Singapore: Springer; In press (accepted September 2020).