UCC

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

THESIS

The role and value of workplace-based assessment in learning

in postgraduate medical education

Aileen Maria Barrett

Thesis submitted for the award of PhD

National University of Ireland, Cork

School of Medicine, College of Medicine and Health

2016

Head of Department: Professor Mary Horgan

PhD Supervisor: Professor Mary Horgan, University College Cork

PhD Co-supervisor: Professor Albert JJA Scherpbier, University of Maastricht

TABLE OF CONTENTS

DECLARATION

I declare that this thesis is my own work and has not been submitted for another degree, either at University College Cork or elsewhere. Where other authors have contributed to the work, this has been explicitly acknowledged and all authors were aware that the work was completed for submission for the award of PhD.

_____   _____

Aileen M. Barrett   Date

ACKNOWLEDGEMENTS

RESEARCH OUTPUTS

Publications

1. Barrett A, Galvin R, Steinert Y, Scherpbier A, O'Shaughnessy A, Horgan M, Horsley T (2015). **Protocol: A BEME (Best Evidence Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees**. *Journal of Systematic Reviews* 4:65 DOI 10.1186/s13643-015-0056-9 (open access)

2. Barrett A, Galvin R, Steinert Y, Scherpbier A, O'Shaughnessy A, Walsh G, Horgan M (2016) **Profiling postgraduate workplace-based assessment implementation; a retrospective cohort study.** *Springer Plus* 5:133 DOI: 10.1186/s40064-016-1748-x (open access)

3. Barrett A, Galvin R, Scherpbier AJJA, Teunissen PW, O'Shaughnessy A, Horgan M. **Is the learning value of workplace-based assessment being realised? A qualitative study of trainer and trainee perceptions and experiences**. *Postgraduate Medical Journal*, 2016; 0:1–5 DOI:10.1136/postgradmedj-2015-133917 (Online first)

Papers in press

1. Barrett A, Galvin R, Steinert Y, Scherpbier A, O'Shaughnessy A, Horgan M, Horsley T. **A BEME review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees**. *Medical Teacher,* Accepted 18th July 2016

Conference presentations

1. Barrett A, Galvin R, O'Shaughnessy A, Tenuissen PM, Scherpbier A, Horgan M. **Fact or fiction? Finding the value of workplace-based in postgraduate medical education.** ASME Conference, Belfast, July 2016 (Poster presentation)

2. Barrett A, Galvin R, O'Shaughnessy A, Tenuissen PM, Scherpbier A, Horgan M. **Reifying the value of workplace-based assessment in postgraduate medical education**. Developing Excellence in Medical Education Conference (DEMEC), Manchester, November 2015 (Poster presentation)

3. Barrett A, Galvin R, O'Shaughnessy A, Tenuissen PM, Scherpbier A, Horgan M. **Fact or fiction? Finding the value of workplace-based in postgraduate medical education.** Queen's University Belfast, SERN Conference, June 2015 (Oral presentation)

4. Barrett A, Galvin R, Steinert Y, Scherpbier A, Horgan M. **Workplace-based Assessment; who is using what and how?** AMEE Conference, Milan, August 2014 (Poster presentation)

Conference presentations (accepted)

1. Barrett A**. PhD Report** – Accepted for presentation at AMEE Conference, August 2017

Grants and awards

1. Scholarly Educational Research Network (SERN), Queen's University Belfast, June 2015. Best Oral Presentation: *Fact or fiction? Finding the value of workplace-based in postgraduate medical education*

2. UCC College of Medicine and Health Doctoral Bursary 2014. €1000 bursary to complete the Qualitative Research Analysis Masterclass at Maastricht University (April 2015)

3. Finalist, UCC Doctoral Showcase, University College Cork, June 2014. Presentation: *Carlsberg don't do feedback for doctors in training but if they did it would be…??*

Invited presentations

1. RCPI Winter Research Evening, February 2014

2. UCC Medical Education Research Group Meeting, May 2015

3. RCPI Winter Research Evening, February 2016

LIST OF ABBREVIATIONS

BEME	Best Evidence Medical and Health Professional Education Collaboration
BST	Basic Specialist Training
CanMEDS	Competency framework of the Royal College of Physicians and Surgeons of Canada
CASP	Critical Appraisal Skills Programme
CBME	Competency-based medical education
CbD	Case-based Discussion
CI	Confidence Interval
COREQ	Consolidated criteria for reporting qualitative research
DOPS	Direct Observation of Procedural Skills
EPAs	Entrustable Professional Activities
ETR	Experience, trajectories, reification
GIM	General Internal Medicine
HST	Higher Specialist Training
Mini-CEX	Mini-Clinical Evaluation Exercise
Mini-PAT	Mini-Peer Assessment Tool
MSF	Multisource feedback (*or* 360° feedback)
OSATS	Objective Structured Assessment of Technical Skills

O&G	Obstetrics and Gynaecology
PRISMA	Preferred reporting items for systematic review and meta-analyses
P-SCO	Pharmacotherapy-Structured Clinical Observation
SD	Standard Deviation
SLEs	Supervised Learning Events
STARLITE	Standards for reporting literature searches (*S*ampling strategy, *T*ype of study, *A*pproaches, *R*ange of years, *L*imits, *I*nclusion and exclusions, *T*erms used, *E*lectronic sources)
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TAB	Team Assessment of Behaviour
RCPI	Royal College of Physicians of Ireland
WBA	Workplace-based Assessment

LIST OF APPENDICES

Chapter 4

Appendix 1   BEME approval letter

Appendix 2   Sample MEDLINE search

Chapter 5

Appendix 1   PRISMA checklist

Appendix 2   STARLITE reporting of literature search

Appendix 3   MEDLINE search

Appendix 4   Database search log

Appendix 5   Kirkpatrick's framework of educational outcomes

Appendix 6   Modified BEME coding sheet

Appendix 7   Weighted kappa calculation

Appendix 8   Voting spreadsheet

Appendix 9   Methodological quality - quantitative studies (BEME criteria)

Appendix 10   Methodological quality qualitative studies (CASP)

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Introduction

Workplace-based assessment (WBA) has been the subject of much debate, contention and varied research evidence since its introduction; a lack of clarity continues to exist as to whether or not the intended learning value of these tools has been realised. While a psychometric focus on the reliability and/or validity of the tools has dominated the literature until recently, contemporary approaches to postgraduate medical education recognise the need for in-depth understanding of the most effective position of these tools within entire programmes of assessment. The purpose of this thesis to establish *what is the role and value of workplace-based assessment in learning in postgraduate medical education?*

Methods

A retrospective cohort study was designed to establish how WBA has been implemented in postgraduate medical education and training programmes in six training bodies and to explore the value of the information contained in the electronic forms. This was followed by a phenomenological study of trainers' and trainees' experiences with WBA and, using a conceptual framework of experience, trajectories and reifications (ETR), how their perceptions of the learning value of WBA have been shaped. The value of WBA in identifying and/or remediating underperformance was explored through the process of a systematic review using the Best Evidence in Medical Education (BEME) review methods.

Results

The cohort study revealed a familiar picture of WBA as a 'tick-box' exercise with limited formative information provided on WBA forms. The *learning* value associated with WBA proved complex to elicit during the semi-structured interviews, however trainees associated WBA with *training* value and a justifiable way to ask for feedback and time with trainers. The BEME review identified 20 studies, including eleven that reported retrospective underperformance data but

could not definitively establish the best conditions under which WBA is effective in identifying or remediating underperformance due to varied methodological approaches, research questions and WBA 'interventions'.

Discussion

Experiences in this study reflect the implementation of a set of individual WBA tools; however the findings suggest that it is the *practice* of formative assessment that is of learning value to trainees. Studies investigating effectiveness of WBA have thus far been limited in their inference of a change in practice as the sole method to evaluate 'evidence' of learning; this study demonstrated that there is a potential learning value in affirmation of good practice and that 'learning' is a complex concept. Determining effectiveness of WBA in identifying or remediating underperformance was limited by lack of high quality prospective studies.

Conclusions

The value of WBA to its users is influenced by experience and is complex to articulate. Observation of practice is essential to the development of the competent physician. WBA appears to have a formative value although this value is complex and difficult to articulate and will require a more considered and explicit definition of learning within this unique educational context. The evidence to date suggests that the value of WBA in learning is associated with narrative feedback; ongoing work is needed to normalise the assessment-feedback process and to influence a positive learning trajectory.

CHAPTER 1

General Introduction

The implementation of workplace-based assessment (WBA) in postgraduate medical education continues to be a source of tension amongst the population for which it was originally intended, that is, the learner (trainee doctors) and teacher (clinical supervisor). Given the widespread, though varied, implementation of this innovation internationally, it appears its acceptability remains limited and its intended purpose – to inform learning through the delivery of effective and constructive observation-based feedback – has not been realised.

In an era of change in postgraduate medical education and moves away from apprenticeship-style, time-based models towards outcome-based education, the integration of robust research-informed workplace assessment practices and methods within programmes of assessment is essential to ensuring that graduates are not only competent to practise, but practise competently.

Since the introduction of workplace-based assessment, the predominant research focus for many years was on the reliability and validity of the individual tools designed to assess performance in the workplace. Research demonstrates that valid and reliable assessments of performance are valued by training institutions and professional and/or regulatory bodies for the purposes of accountability, public reassurance and regulation. Evidence also indicates that under certain conditions, these tools can reliably measure certain performance metrics. However, in the midst of this psychometric wave of research, confusion has continued among users as to the purpose and aim of these assessments and we have not yet ascertained whether WBA has a learning value as intended.

By exploring the learning value of WBA to its intended users, this will facilitate the (re)consideration of the current methods of implementation of WBA and assist in finding its most appropriate and effective position within a programme of assessment. While we have explored this problem within a specific geographical

and disciplinary context the aim of the research is to contribute to the wider international conversation on formative assessment in postgraduate medical education.

In this chapter I will outline the problems identified as a basis for this work, the questions that guided the programme of research and the methodological and conceptual approaches to the individual studies contained within research programme.

Formative assessment in postgraduate medical education

Formative assessment in postgraduate medical education has taken the shape of workplace-based assessment, a suite of tools designed to solve two problems; those of poor supervision and poor feedback practices. The first such tool, the mini-clinical evaluation exercise (mini-CEX) was designed in response to research that trainees were rarely being supervised in clinical practice[1] with the intention of informing learning by providing a structure for constructive, observation-based feedback.[2, 3]

The design of the WBA tools was rooted in emerging educational research on the concept of assessment-for-learning,[4-6] in which assessment is used both diagnostically (to ascertain the learner's current level of knowledge or skills) and to inform the next learning objectives or activities. WBA also satisfied the highest point of Miller's pyramid, i.e. assessments of what the learner 'does' in practice.[7] These low-stakes assessments were not intended to be used to make high-stakes decisions e.g. progression from one year to the next; rather there was an emphasis on the narrative feedback provided and the conversation that should theoretically follow from the assessment.

Over the last twenty years since the mini-CEX was introduced, more than fifty individual tools have been designed, researched and implemented across the world to varying effect.[8-11] Some of these tools have become mainstream in many postgraduate training programmes including the mini-CEX, Direct Observation of

Procedural Skills (DOPS), case-based discussion, Objective Structured Assessment of Technical Skills (OSATS) and 360⁰ or multisource feedback (MSF). Each tool is designed to provide the trainee with feedback on their performance in a specific aspect of practice including, for example communication skills in delivering a diagnosis, clinical skills in physical examination and clinical reasoning.

While these tools are familiar in name to most trainees, their actual implementation and individual formats have varied significantly. In general, trainees take responsibility for requesting or initiating assessments and ensuring that these are uploaded or recorded to their training portfolios or records.

Some programmes have implemented WBA tools using narrative rating scales (e.g. above expectation), global rating scales (e.g. 'good', 'very good') or general feedback only; others have assigned numerical rating scores[8, 12] or grades.[13] The shift to include grades or numerical ratings reflects to some extent the era in which these assessments were introduced; the predominant psychometric approach to evaluating assessment tools led to a large literature on the reliability and validity of individual tools.[14] This approach was also driven by recognition of the need for better accountability on the part of training bodies and programmes to ensure that graduates were fit to practise and therefore assessed effectively during their training.[15]

Innovation in the changing landscape of postgraduate medical education

Three other significant changes have also occurred over the last twenty years that have had an impact on the implementation of workplace-based assessment. The first and most significant global change in medical education has been the widespread interest in and adoption of outcome-based medical education, a move away from the traditional time-based 'teabag-steeping' model[16] to one that acknowledges the need to articulate and describe the skills and abilities required of a practising clinician, and also recognises that not all trainees may develop those skills and abilities within this randomly-allocated timeframe.[17, 18] Outcome-based

education and training is centred on the need for regular, ongoing and informative workplace assessment through the use of 'entrustable professional activities' (EPAs) against which the need for supervision, feedback and intervention decisions can be judged.[19-21] Their success will be determined by the integration of assessment methods and practices that are fit-for-purpose and longitudinally aligned to the development of competencies and skills.[17, 22]

A significant body of work is also emerging within this context around the development of programmatic approaches to assessment; the recognition that both summative and formative assessments - or 'high-stakes' and 'low-stakes' respectively - all have a place in the overall assessment of trainee learning, but that different methods are required to assess or evaluate different skills and abilities.[19, 23-27] In a recent review of formative assessment research, Paul Black has proposed that in resolving the issues and tensions in the relationship between formative and summative assessments, the approach of 'marriage counselling rather than divorce' should be adopted.[28] In the context of medical education, this will require a better understanding of the role and position of individual workplace-based assessments, and whether a series of low-stakes assessments can or should inform high-stakes (e.g. progression) decisions.

Secondly, the research community has begun to embrace qualitative and mixed methods research,[29] allowing access to richer and deeper information on the acceptability of such innovations[30, 31] and also pushing the boundaries of our understanding of concepts that are clearly defined in medicine (such as 'validity') but which are not as easily transferable to medical education.[12, 32-35] Using this specific example, validity is defined within a clinical context as the ability of a scale or tool to measure that which it purports to be able to measure. A number of studies have attempted to use this psychometric approach to workplace-based assessments with limited translation.[12, 33, 34] However, recently there has been a recognition that if a workplace-based assessment tool is designed to inform learning, then the evaluation of the effectiveness of the tool in achieving this aim needs to be considered in different and arguably more complex way.[36]

Thirdly, since the development of WBA, educational research has also continued to address the issue of what exactly *is* formative assessment and how does it impact on learning? A number of researchers have attempted to define formative assessment and assessment-for-learning; however a consensus on these definitions remains elusive.[37, 38] In the absence of this consensus, most definitions refer to the almost retrospective function of the assessment i.e. it is formative if it has contributed to or further informed learning.[37] In a large-scale review of theory of formative assessment, Clark (2012) defines assessment-for-learning in the context of self-regulated learning, in which a number of social cognitive and sociocultural theories are used to explain how learners can become self-aware and that one aspect critical to the development of this skill is that of formative assessment.[39]

The ongoing issues of acceptability and implementation of formative assessment in primary, secondary and third-level education mirror those encountered by workplace-based assessment in postgraduate medical education. Despite many attempts to communicate the purpose of WBA as formative, there is still widespread confusion among many training populations as to the purpose of the assessments.[11, 19, 40, 41] For example, where individual assessments might not be used to make progression judgments, the inclusion of minimum WBA requirements for a training year or specific procedure have led to an understanding of these assessments as summative.[19]

Research demonstrates that trainees and trainers continue to view the assessments as time-consuming and of limited relevance to postgraduate training outcomes, curricula or individual development particularly in the context of busy clinical workload.[11] Evidence also indicates that feedback delivery[42] and feedback reception[43-45] continue to be a source of tension among and between trainers and trainees. We know that implementing requirements of up to ten WBAs to ensure reliability and validity[46-48] has become less feasible – and less acceptable - in increasingly busy, pressured and changing clinical environments particularly where these are not perceived to be linked to curriculum or programme outcomes.[19, 40] We also know that WBA tools do not perform well as summative assessments, either in real-time or simulated settings[12, 49] as they are subject to confounding

issues of variability in rater judgment and differences[50, 51] yet they continue to be required as stand-alone components of many training portfolios.

With the widespread adoption of outcome-based education models and the newer research methods and practices available to the medical education community, we are now in an era where it is possible evaluate these assessments from multiple perspectives, including the impact of trainee performance on the quality of patient care.[52]

So the question now arises as to whether or not WBA, in its current form, has a role and value in postgraduate education and training, or whether with modifications suggested by Massie and Ali[11] and with a critical understanding of the learning value as originally intended, we can better position these assessments within a programme of assessment to maximise trainee learning and development and align learning outcomes to training programme curricula.

The research aim

The aim of this thesis is to ask the question *what is the value of workplace-based assessment in learning in postgraduate medical education*?

In asking complex questions, we recognised that one research method or approach would not be sufficient to find complex answers. We also recognised that local implementation of WBA tools and methods varies widely and we could not assume that implementation in Ireland had mimicked that of anywhere else. However, the format and content of the tools adopted by the institution were consistent with those in use internationally and included the mini-CEX, DOPS, case-based discussion and OSATS (appendices 1-4) and use ratings described by 'expectations' i.e. '*below expectation*', '*well above expectation*'.

We therefore started by asking what was happening in our own context in the three years since the introduction of WBA. *Chapter 2* details the retrospective cohort study used as a scene-setting study for the programme of research. In this study we

evaluated the quality of the content of submitted workplace-based assessments over a 12-month period including the presence/absence of feedback and identification of components deemed 'borderline' or 'below expectation'. We also looked at the numbers of WBAs completed by trainees and the timing of these WBAs in the training year.

The second study (*Chapter 3*) set out to explore potential reasons for the patterns of implementation we had uncovered. In this qualitative study, we asked trainers and trainees to share their perceptions of the learning value of WBA and we attempted to illuminate how their individual and collective WBA experiences had shaped these perceptions.

In the final study (*Chapter 4 and 5*) we attempted to explore the value of workplace-based assessment in identifying a specific cohort of trainees, those who are poorly or underperforming. One of the main tenets of formative assessment is that of diagnostic assessment – establishing a learner's current level to inform learning strategies and activities that move the learner from their current to the desired level of performance, knowledge or skills. While emerging research on the impact of WBA on observable changes in practice has uncovered little supporting evidence, most of this literature to date has focused on general cohorts of trainees, many of whom may not have required changes in practice, but for whom 'learning' may have been an affirmation of current practice. We aimed to address this gap in the literature by exploring the use of WBA specifically in identifying and/or remediating poor performance. This study, undertaken as BEME review, allowed us to consolidate the research evidence to date and to explore the totality of the evidence on the use of WBA in this cohort.

Chapter 6 brings together the findings from these studies to answer our original question and attempts to fill the gaps in the literature that could impact on the implementation of WBA in a constructive and effective way as we move towards programmes of assessment in outcome-based education models. The chapter ends with a discussion on the role and position of this medical education innovation now and into the future and by proposing a reconceptualisation of WBA not simply as a

set of tools, but as a *practice* supported by well-designed tools. This chapter

concludes with recommendations for practice and future research.

Positioning the researcher; a reflexive perspective

This programme of research originated in my desire to solve a problem; how can we best facilitate the delivery of effective feedback to postgraduate medical trainees and is WBA the best way to do this?

My professional role involves the integration of medical education innovations into postgraduate training curricula. One of my first tasks within this current role was to evaluate the implementation of the new assessment methods - including WBA - implemented by the Royal College of Physicians of Ireland.

As I was not involved in the original implementation plans and decisions, this research has been conducted from a position somewhere between that of 'insider' or 'practitioner-researcher'[53] and an 'outsider'; on the one hand attempting to justify implementation decisions that had been made prior to my appointment and on the other, often struggling with the impact of those decisions on some aspects of postgraduate medical education and training. I am also an outsider in terms of my background as a clinician but not a doctor, and occasionally battling institutional perceptions of the role of educationalists in postgraduate medical education.

One of the main tensions that I considered in the development of my approach to the programme of research was that between my own paradigmatic preferences towards social-constructivism and the manner in which WBA had been implemented. Implementation of mandatory minimum requirements - e.g. for two mini-CEX assessments a year - conflicted significantly with my view that formative assessment and feedback needed to be constructed within the context of the setting and not as a 'truth' or 'accuracy' in performance assessment. In order to somehow manage these (mostly internal) tensions, I devised a programme of research that looked at WBA implementation in a 360-degree way, incorporating factual, observation-based research approaches with a qualitative study looking at the value of this innovation to its main users i.e. trainers and trainees.

I also realised that in order to be able to answer questions about what works for whom and why, I needed to approach this question from a systematic research

base to which the institution could relate. In 2006 Albert et al[54] proposed a better alignment between medical education research and practice by using Bourdieu's concept of the 'field', in which each field is bounded by two 'poles', a 'pole of production for users' and a 'pole of production for producers'. Those researchers positioned within the former pole tend to consider the value or implications of the research outputs to the end users, the 'so what?' factor, while those in the latter pole value more highly the contribution to knowledge and academic excellence in the research itself. In this programme of research, my aim was to somehow bridge both; by developing my own research skills and competence across a range of methods, I hope to be able to credibly and legitimately have a voice in the international conversation on WBA which will allow me to continue to contribute towards knowledge enhancement and to producing research that is of value and of practical relevance to postgraduate medical education.

In all of the studies in this programme of research I have endeavoured to consider, acknowledge and communicate the influence of my role and perspectives on the choice of research question, study methods and research outcomes. This includes recognition of my desire to solve a problem and improve training processes and methods, particularly for trainees and ultimately, to the benefit of patients. However, instead of attempting to somehow overcome or 'bracket' those assumptions and perspectives, I have endeavoured to ensure that they are instead a transparent component in my interpretation and in the validity of the work completed.

References

[1]   Day S, Grosso L, Norcini J, Blank L, Swanson D, Horne M. Residents'
perception of evaluation procedures used by their training program. *Journal of
General Internal Medicine*. 1990;**5**:421-426.

[2]   Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (Clinical Evaluation
Exercise): A Preliminary Investigation. *Annals of Internal Medicine*. 1995;**123**:795-
799.

[3]   Norcini J, Burch V. Workplace-based assessment as an educational tool:
AMEE Guide No. 31. *Medical Teacher*. 2007;**29**:855-871.

[4]   Black P, Wiliam D. *Inside the black box; Raising standards through classroom
assessment*. London GL Assessment; 1998(b).

[5]   Black P, Wiliam D. Assessment and Classroom Learning. *Assessment in
Education: Principles, Policy & Practice*. 1998a;**5**:7-74.

[6]   Wiliam D. Formative Assessment: Getting the Focus Right. *Educational
Assessment*. 2006;**11**:283-289.

[7]   Norcini JJ. Work based assessment. *BMJ : British Medical Journal*.
2003;**326**:753-755.

[8]   Kogan JR, Holmboe ES, Hauer KE. Tools for Direct Observation and
Assessment of Clinical Skills of Medical Trainees. *JAMA: The Journal of the American
Medical Association*. 2009;**302**:1316-1326.

[9]   Miller A, Archer J. Impact of workplace based assessment on doctors'
education and performance: a systematic review. *BMJ*. 2010;**341**.

[10] Overeem K, Wollersheim H, Driessen E, et al. Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study. *Medical Education*. 2009;**43**:874-882.

[11] Massie J, Ali J. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Advances in Health Sciences Education*. 2015:1-19.

[12] Hatala R, Cook D, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Advances in Health Sciences Education*. 2015:1-27.

[13] Lefroy J, Hawarden A, Gay SP, McKinley RK, Cleland J. Grades in formative workplace-based assessment: a study of what works for whom and why. *Medical Education*. 2015;**49**:307-320.

[14] Gormley GJ, Johnston JL. 'He'll make a good doctor': examiners' little black books and the assessment research agenda. *Medical Education*. 2015;**49**:138-139.

[15] GMC. Workplace based assessment: A guide for implementation. General Medical Council, 2010.

[16] Hodges B, Lingard L. Introduction. In: Hodges BD, Lingard L, eds. *The Question of Competence; Reconsidering Medical Education in the Twenty-First Century*. ILR Press, Ithaca and London, 2014.

[17] Frank J, Snell L, Ten Cate O, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;**32**:638 - 645.

[18] Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Medical Teacher*. 2010;**32**:651-656.

[19]   Bok H, Teunissen P, Favier R, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education*. 2013;**13**:123.

[20]   Hauer KE, Soni K, Cornett P, et al. Developing entrustable professional activities as the basis for assessment of competence in an internal medicine residency: a feasibility study. *Journal of General Internal Medicine*. 2013;**28**:1110-1114.

[21]   ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Medical Teacher*. 2015;**37**:983-1002.

[22]   Holmboe ES. Realizing the Promise of Competency-Based Medical Education. *Academic Medicine*. 2015;**90**:411-413.

[23]   Van der Vleuten C, Schuwirth L. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;**39**:309 - 317.

[24]   Dijkstra J, Van der Vleuten C, Schuwirth L. A new framework for designing programmes of assessment. *Adv Health Sci Educ*. 2009;**15**:379 - 393.

[25]   Schuwirth L, Van der Vleuten C. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;**33**:478 - 485.

[26]   van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Medical Teacher*. 2012;**34**:205-214.

[27]   Heeneman S, Oudkerk Pool A, Schuwirth LWT, van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Medical Education*. 2015;**49**:487-498.

[28]   Black P. Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*. 2015;**22**:161-177.

[29]   Hodges B. Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical Teacher*. 2013;**35**:564-568.

[30]   Fokkema JPI, Teunissen PW, Westerman M, et al. Exploration of perceived effects of innovations in postgraduate medical education. *Medical Education*. 2013;**47**:271-281.

[31]   Fokkema JP, Scheele F, Westerman M, et al. Perceived effects of innovations in postgraduate medical education: a Q study focusing on workplace-based assessment. *Academic Medicine*. 2014;**89**:1259-1266.

[32]   Cook D. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Advances in Health Sciences Education*. 2014:1-6.

[33]   Cook D, Zendejas B, Hamstra S, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education*. 2014;**19**:233-250.

[34]   Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 2015;**49**:560-575.

[35]   St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Medical Education*. 2015;**49**:548-550.

[36]   Govaerts M. Workplace-Based Assessment and Assessment for Learning: Threats to Validity. *Journal of Graduate Medical Education*. 2015;**7**:265-267.

[37]   Wiliam D. What is assessment for learning? *Studies In Educational Evaluation*. 2011;**37**:3-14.

[38]   Bennett RE. Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*. 2011;**18**:5-25.

[39]   Clark I. Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*. 2012;**24**:205-249.

[40]   Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014;**4**.

[41]   Bok HGJ. Competency-based veterinary education: an integrative approach to learning and assessment in the clinical workplace. *Perspectives on Medical Education*. 2015;**4**:86-89.

[42]   Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*. 2012;**46**:201-215.

[43]   Watling C. Cognition, culture, and credibility: deconstructing feedback in medical education. *Perspectives on Medical Education*. 2014;**3**:124-128.

[44]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education*. 2013;**47**:585-594.

[45]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: The potential and limitations of regulatory focus theory. *Medical Education*. 2012;**46**:593-603.

[46]   Alves de Lima A, Conde D, Costabel J, Corso J, Vleuten C. A laboratory study on the reliability estimations of the mini-CEX. *Advances in Health Sciences Education*. 2013;**18**:5-13.

[47]   Cook DA, Beckman TJ, Mandrekar JN, Pankratz VS. Internal structure of mini-CEX scores for internal medicine residents: Factor analysis and generalizability. *Advances in Health Sciences Education*. 2010;**15**:633-645.

[48]   Weston PSJ, Smith CA. The use of mini-CEX in UK foundation training six years following its introduction: Lessons still to be learned and the benefit of formal teaching regarding its utility. *Medical Teacher*. 2014;**36**:155-163.

[49]   Setna Z, Jha V, Boursicot KAM, Roberts TE. Evaluating the utility of workplace-based assessment tools for speciality training. *Best Practice & Research Clinical Obstetrics & Gynaecology*. 2010;**24**:767-782.

[50]   Govaerts M, Schuwirth L, Van der Vleuten C, Muijtjens A. Workplace-based assessment: effects of rater expertise. *Advances in Health Sciences Education*. 2011;**16**:151-165.

[51]   Gingerich A, Regehr G, Eva KW. Rater-Based Assessments as Social Judgments: Rethinking the Etiology of Rater Errors. *Academic Medicine*. 2011;**86**:S1-S7.

[52]   Carraccio C, Englander R, Holmboe ES, Kogan JR. Driving Care Quality: Aligning Trainee Assessment and Supervision Through Practical Application of Entrustable Professional Activities, Competencies, and Milestones. *Academic Medicine*. 2015;**Publish Ahead of Print**.

[53]   Dhalberg L, McCaig C. Practitioner research and the research process. In: Dahlberg L, McCaig C, eds. *Practical Research and Evaluation A start-to-finish guide for practitioners*. SAGE Publications Ltd, London, 2010; 1-12.

[54]   Albert M, Hodges B, Regehr G. Research in Medical Education: Balancing Service and Science*. *Advances in Health Sciences Education*. 2006;**12**:103-115.

Profiling postgraduate workplace-based assessment implementation in Ireland; a retrospective cohort study

Authors

Aileen Barrett, Rose Galvin, Yvonne Steinert, Albert Scherpbier, Ann O'Shaughnessy, Gillian Walsh, Mary Horgan

Abstract

Background

In 2010, workplace-based assessment (WBA) was formally integrated as a method of formative trainee assessment into 29 basic and higher specialist medical training (BST/HST) programmes in six postgraduate training bodies in Ireland. The aim of this study is to explore how WBA is being implemented and to examine if WBA is being used formatively as originally intended.

Methods

A retrospective cohort study was conducted and approved by the institution's Research Ethics Committee. A profile of WBA requirements was obtained from 29 training programme curricula. A data extraction tool was developed to extract anonymous data, including written feedback and timing of assessments, from Year 1 and 2 trainee ePortfolios in 2012-2013. Data were independently quality assessed and compared to the reference standard number of assessments mandated annually where relevant.

Results

All 29 training programmes mandated the inclusion of at least one case-based discussion (max =5; range: 1-5). All except two non-clinical programmes (93%) required at least 2 mini-Clinical Evaluation Exercise assessments per year and Direct Observation of Procedural Skills assessments were mandated in 27 training programmes over the course of the programme.

WBA data were extracted from 50% of randomly selected BST ePortfolios in four programmes (n=142) and 70% of HST ePortfolios (n=115) in 21 programmes registered for 2012-2013. Four programmes did not have an eligible trainee for that academic year. In total, 1142 WBAs were analysed.

A total of 164 trainees (63.8%) had completed at least one WBA. The average number of WBAs completed by HST trainees was 7.75 (SD 5.8; 95%CI: 6.5-8.9; range 1-34). BST trainees completed an average of 6.1 assessments (SD 9.3; 95% CI: 4.01-8.19; range 1-76). Feedback – of varied length and quality - was provided on 44.9% of assessments. The majority of WBAs were completed in the second half of the year.

Conclusion

There is significant heterogeneity with respect to the frequency and quality of feedback provided during WBAs. The completion of WBAs later in the year may limit available time for feedback, performance improvement and re-evaluation. This study sets the scene for further work to explore the value of formative assessment in postgraduate medical education.

Keywords

Workplace-based assessment; postgraduate medical education; retrospective cohort study

Introduction

Workplace-based assessment (WBA) was originally mooted as a formative – or 'assessment-*for*-learning' – practice with a primary aim of impacting trainee learning and development and to assist in focusing the trainee's learning plans.[1] The format of the assessment takes place in real time, with the supervisor observing the trainee in a specific aspect of clinical practice. Since its introduction many tools have been developed[2] to structure feedback on specific aspects of a trainee's performance.

Over time, the use of WBA has expanded to include a quality assurance role[3] and has been mooted as a method of early identification of poor performance.[4] Implementation of WBA internationally has met with varied levels of success and acceptability[5] with many ongoing reservations regarding the practical feasibility of performing multiple assessments in order to comply with recommendations for good reliability while attempting to maintain the formative function of these assessments.[6] The introduction of what is viewed as an additional demand on trainer and trainee time, in an increasingly busy and unstructured environment has also impacted on the acceptability of these learning 'innovations'.[5, 7]
One of the main criticisms of the implementation of WBAs has emerged where the assessments are not mapped to training programme outcomes or aligned with a defined programme of assessment throughout training.[8] Poor communication of the formative purpose of WBA has also emerged as a critical barrier to successful implementation of these tools.[6] Attempts to communicate the formative nature of the assessments in the UK by changing the name to 'supervised learning events' have also been met with mixed opinions.[9]

The focus of workplace-based assessment research has, however, begun to take a new direction. While acknowledging the limitations of workplace-based assessment as individual summative judgments of performance, the place of these tools within a programme of assessment hinges more on their validity as formative assessments, than their reliability as summative assessments.[10-13] The role of

narrative feedback in this conceptualisation of validity becomes therefore increasingly important.

In the Irish context, WBA was introduced as mandatory component of postgraduate medical training across six training bodies in 2010. The mini-clinical evaluation exercise (Mini-CEX;) and case-based discussion (CbD;) were included across all disciplines while the Direct Observation of Procedural Skills (DOPS;) assessment was included for disciplines with procedural skill requirements. The OSATS (Objective Structured Assessment of Technical Skills;) - with procedure-specific adaptations - was implemented in both basic and higher specialist training programmes in Obstetrics and Gynaecology. Procedure-specific DOPS forms were also developed and implemented for higher specialist training in gastroenterology (Appendices 1-3).

Research aim

The research question posed by this study is 'how have workplace-based assessments been integrated into higher specialist training programmes in medicine in Ireland?'

The study comprised three key objectives:

1. to describe the level of implementation of WBA in postgraduate Basic Specialist Training (BST) and Higher Specialist Training (HST) programmes in one postgraduate medical training institution in Ireland
2. to compare the findings with those published from other training jurisdictions
3. to explore the quality of written feedback provided in these assessments

Conceptual framework

This study was guided by work in two key areas of educational research, formative assessment theory[14, 15] and guidelines for good practice in effective feedback.[16, 17] Contemporary formative assessment theory proposes that all assessment should guide learning and development.[18, 19] Guidelines for good practice suggest that in order to be effective, feedback must be, among other factors, specific, timely and

result in a further plan for development.[16] The mechanisms by which feedback can be deemed to be successful in this purpose remain challenging to elucidate and the learner's response to that feedback - and therefore its ultimate use - is less predictable.[20-22] This study therefore only addressed evidence of feedback provided on written assessments and did not attempt to link this directly to evidence of learning.

Methods

Study design

This study was conducted using a retrospective cohort design. The STROBE standardised reporting guidelines were followed to ensure the standardised conduct and reporting of the research.[23, 24] Ethical approval was obtained from the institution's Research Ethics Committee (Appendix 4)

Setting and study size

The study was conducted over a 3-month period from September-December. Data were extracted anonymously from trainee ePortfolios for the academic year 2012-2013 (July-July). In 2011 a new ePortfolio replaced an existing paper-based recording system for trainees commencing programmes in that year. Therefore only data for Year I and Year II trainees (BST and HST) was available to access for this study. In order to obtain a truly representative picture of the level of implementation of WBA, and considering the small total population size, 50% of registered BST ePortfolios and 70% of HST ePortfolios were included in the study.

Data extraction

A data extraction tool was developed to extract anonymous data from trainee ePortfolios prior to the study commencement. This tool (Fig 1) was designed to extract data on key 'quality indicators' of effective feedback, adapted from a number of sources including Nicol and MacFarlane-Dick's 'seven principles of good feedback practice'[16] and the WBA form content in use on these assessments. These indicators were assessed as binary outcomes (present/absent) and included the presence of learner-centred feedback specific to the assessment, learning goals and

further follow-up where any competence was deemed to be 'borderline' or 'below expectation'. The tool was piloted using data from five sample ePortfolios with one minor change to the use of 'weeks' instead of months in ascertaining the timing of the assessment completion. The timing of WBAs was therefore measured in weeks from the start of the academic year (9th July 2012).

Quality check

Data were extracted by the principal investigator (AB) and a quality check of 10% of the data extraction sheets was conducted by a second author (RG) prior to analysis. No extraction errors were identified; however it was agreed by the two authors to exclude three trainees' data from the final analysis due to completion errors identified in those ePortfolios.

Data analysis

The profile of WBA requirements was analysed descriptively from an Excel spreadsheet as were data extracted from ePortfolios. Binary data are presented as proportions where the denominator represents the total number of assessments completed in the programme. Summary means and standard deviations (SDs) are reported for continuous data, with corresponding 95% confidence intervals (CI). Ranges are reported to illustrate the spread in the data. Data were compared to the reference standard number of assessments mandated annually where relevant.

Results

Data were extracted from a random selection 50% of BST ePortfolios in four programmes (n=142) and 70% of HST ePortfolios (n=115) in 21 programmes registered for 2012-2013. Four programmes did not have an eligible trainee for that academic year. A total of 1142 individual assessments were analysed.

WBA programme integration profile

All 29 programme curricula mandated at least one CbD annually (range 1-5). Annual mini-CEX assessments were required in all but two non-clinical specialties (range 1-

4). DOPS requirements varied from 0-37 and most were required over the course of the training programme to allow for variations in opportunities to develop procedural skills in individual rotations. Two 'non-procedural' programmes did not have any DOPS requirement.

In HST, General Internal Medicine (GIM) training is completed alongside one of eight subspecialties. Trainees in these programmes complete at least one year of 'high intensity GIM' in which they must complete GIM curriculum requirements only and a 'non-GIM' year in which they complete their specialty requirements. For all other years, trainees complete requirements for both their GIM and specialty curriculum.

WBA completion profile

The majority of trainee ePortfolios (164; 63.8%) contained at least one completed WBA (76.5% HST; 53.5% BST). The average number of WBAs completed by individual HST trainees was 7.75 (SD 5.8; 95%CI: 6.5-8.9; range 1-34). BST trainees completed an average of 6.1 assessments (SD 9.3; 95% CI: 4.01-8.19; range 1-76).

The 'quality indicators' for each WBA are detailed in Tables 1 and 2.

Assessments were mostly completed in the second half of the training year, after week 30. Trainees were more likely to complete DOPS/OSATS than Mini-CEX or CbD assessments (ratio 3:1); 76 BST trainees completed 281 DOPS/OSATS, 88 Mini-CEX and 94 CbD assessments. A similar pattern emerged at HST where 88 trainees completed 359 DOPS/OSATS, 153 Mini-CEX and 167 CbD assessments. There were many errors in ePortfolio completion among 'dual' specialty trainees with WBAs entered into the incorrect logbook or use of the same WBA in both.

Table 1: Basic Specialist Training Results

	DOPS/OSATS	MINI-CEX	CBD
TOTAL NO COMPLETED	281	88	94
AVERAGE NO. WEEKS IN POST BEFORE WBA COMPLETED	31.7 (95%CI: 30.1-33.3; SD 13.6; range 3-52)	35.2 (95%CI: 32.4-38.1; SD 13.5; range 7-52)	34.4 (95%CI: 31.4-37.3; SD 14.5; range 5-52)
ENTRIES DEMONSTRATING DEFINED GOALS	0	1 1.13%	0
ENTRIES WITH EVIDENCE OF FEEDBACK	174 61.9%	54 (61.3%)	33 (35.1%)
EVIDENCE OF ANY COMPETENCE 'BORDERLINE' OR 'BELOW EXPECTATION'	38 (13.5%)	1 1.13%	1 1.06%
EVIDENCE OF FOLLOW-UP	17 (44.7%)	0	0

Feedback was provided on 44.9% of assessments; however the content of this feedback varied from one word (e.g. excellent) to complete sentences about the assessment episode. Trainer comments that pertained to the case (e.g. 'complex case') were not included as feedback in the analysis.

A total of 40 BST WBAs (8.63%) and 12 HST WBAs (1.76%) extracted contained a competence or component that was 'borderline' or 'below expectation'. Of the 38 BST DOPS/OSATS assessments with a component deemed to be 'borderline' or below expectation, all were from within one speciality and 17 (44.7%) were followed up with a second WBA in the same procedure. The 10 HST DOPS with a component identified as 'borderline' or 'below expectation' were also from the same specialty; however none of these ePortfolios demonstrated evidence of follow-up.

Table 2: Higher Specialist Training Results

	DOPS/OSATS	Mini-CEX	CbD
Total no completed	359	153	167
Average no. weeks in post before WBA completed	30.1 (95% CI: 28.6-31.5; SD 14.1; range 1-52)	33.6 (95%CI: 31.6-35.6' SD 13.7; range 3-52)	32.6 (95%CI: 31.7 -35.5; SD 12.4; range 2-52)
Entries demonstrating defined goals	1 0.27%	1 0.65%	1 0.59%
Entries with evidence of feedback	104 (28.9%)	102 (66.6%)	46 (27.5%)
Evidence of any competence 'borderline' or 'below expectation'	10 (2.78%)	0	2 (1.19%)
Evidence of follow-up	0	n/a	0

Discussion

The aim of this study was to determine the patterns of workplace-based assessment integration throughout postgraduate medical training curricula in six training bodies. Our main findings demonstrate that while the level of implementation has been varied, the majority of trainees have experienced at least one WBA during the academic year.

The picture that has emerged in this observational study compares in many ways with the issues identified internationally; particularly those related to ineffective feedback and limited formative impact. We identified that the *documentation* of effective written feedback was limited; however, as these assessments take place in real-time with the trainer and trainee present, *verbal* feedback, which is not then transferred to the assessment forms, may also take place. A number of international institutions have implemented WBA smart-phone and tablet 'apps' which allow for real-time completion and uploading of the assessment feedback. Another barrier to the provision of feedback in our study may have been the lack of an explicitly-titled free-text '*feedback*' section ; on these assessments the free text section was titled '*comments*' and therefore was interpreted by some trainers as comments on the case, not on the trainee performance.

In our study, both at BST and HST level, trainees were more likely to complete DOPS assessments than the mini-CEX or CbD. This finding is in keeping with a UK study of dermatology trainees where the authors reported that 138 trainees completed 251 DOPS compared with 142 mini-CEX assessments.[4] In this study respondents reported that the Mini-CEX and Multisource Feedback (MSF) tended to feel more 'artificial' than DOPS; they also reported dissatisfaction with the quality of feedback provided on all assessments, despite an overall positivity about the benefits of WBAs. While there is limited empirical research exploring trainer and trainee preferences regarding assessment, it may be that trainers and trainees perceive DOPS as a more objective measure of performance as opposed to the more subjectively-perceived assessments of, for example, communication and professionalism. However, it is interesting to note that in a 2009 study of psychiatry

trainees – for whom procedure-based WBAs are not usually required – Menon et al[25] also reported that trainees were 'unimpressed' with the introduction of these assessments, querying their reliability, validity and impact on the quality of training.

Our study found that the majority of WBAs took place in the second half of the year. This pattern, along with the limited provision of written feedback and follow-up assessments, appears to point towards a limited use of these assessments to inform learning and development. During the implementation of WBAs in the UK, one 2011 study of paediatric trainees[26] reported that WBAs were still viewed as a 'tick-box' exercise. Menon et al reported that psychiatry trainers[27] and trainees[25] understood that the introduction of WBAs was both driven by a desire to improve training but that it was also 'politically driven'; comments from these trainees also referenced the 'tick-box exercise' designed purely to fulfil end-of-year assessment requirements. In a recent review of the issues underlying the problems encountered in WBA implementation Swayamprakasam et al[28] also pointed towards the need for widespread communication strategies to inform – or re-inform – the understanding of the purpose of WBA.

The potential 'floor' and 'ceiling' effect of WBA also warrants further investigation. In this study, the low number of assessments documenting a competence that was 'borderline' or 'below expectation' raises a number of issues around 'failure to fail'. The reluctance and anxiety of trainers around the delivery of negative feedback is well documented[29] as are issues with the rating systems used to structure this feedback.[30] In our assessments, the use of an 'expectations' rating system (i.e. 'above expectation', 'meets expectations') in Mini-CEX and CbD assessments, without explicit reference to curriculum outcomes or competencies, may also have been perceived as overly-subjective and less conducive to learning.

This is the first large-scale study of WBA implementation in Ireland. The methodology employed to conduct the study was rigorous and quality checks were implemented to ensure the quality and accuracy of the data. The study provides and overview of the varied integration of the assessments since the introduction of

the tools and has highlighted similar issues to those identified internationally. The study was designed to provide a thorough background in developing an extensive programme of research on WBA in the Irish postgraduate medical education context and will form the basis of a large in-depth qualitative study to explore the value of WBAs to both trainers and trainees. The findings have also highlighted a number of areas for further development of the assessment, particularly regarding the implementation and assessment of same. One of the main limitations of the study lies in the evaluation of the quality of feedback; only written feedback was extracted which may not accurately or fully reflect the quality or richness of verbal feedback provided at the end of the workplace-based assessments.

Conclusion

This study was developed as a 'scene-setting' exploration of what has happened within our medical training programmes at our institution since the introduction of workplace-based assessments in 2010; however it reflects and adds to the international body of work on workplace-based assessment implementation. As is the case internationally, issues persist in the successful implementation of formative assessment in postgraduate medical education. Recommendations based on this study and a subsequent larger qualitative study, are currently in motion with the aim of further contributing to the international discussion on the value of formative assessment in trainee development.

Authors' Contributions

AB developed the study concept. AB and RG and carried out data extraction and analysis. AB and RG drafted manuscript. RG, MH, AO'S, GW, YS and AS provided guidance throughout the study and provided write-up assistance. All authors read and approved the final manuscript.

References

[1]   Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (Clinical Evaluation Exercise): A Preliminary Investigation. *Annals of Internal Medicine*. 1995;**123**:795-799.

[2]   Kogan JR, Holmboe ES, Hauer KE. Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees. *JAMA: The Journal of the American Medical Association*. 2009;**302**:1316-1326.

[3]   Black D, Welch J. The under-performing trainee – concerns and challenges for medical educators. *Clin Teach*. 2009;**6**:79-82.

[4]   Cohen S, Farrant P, Taibjee S. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol*. 2009;**16**:34 - 39.

[5]   Fokkema JPI, Teunissen PW, Westerman M, et al. Exploration of perceived effects of innovations in postgraduate medical education. *Medical Education*. 2013;**47**:271-281.

[6]   Bok H, Teunissen P, Favier R, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education*. 2013;**13**:123.

[7]   Fokkema J, Teunissen PW. Assessing the assessment of interventions: we're not there yet. *Medical Education*. 2013;**47**:954-956.

[8]   Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. *Medical Teacher*. 2013;**35**:569-574.

[9]   Ali J. Workplace-based assessments: lost in translation? *The Clinical Teacher*. 2014;**11**:68-69.

[10]   Cook D, Zendejas B, Hamstra S, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education*. 2014;**19**:233-250.

[11]   Hatala R, Cook D, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Advances in Health Sciences Education*. 2015:1-27.

[12]   Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 2015;**49**:560-575.

[13]   St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Medical Education*. 2015;**49**:548-550.

[14]   Clark I. Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*. 2012;**24**:205-249.

[15]   Bennett RE. Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*. 2011;**18**:5-25.

[16]   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**:199-218.

[17]   Watling C. Cognition, culture, and credibility: deconstructing feedback in medical education. *Perspectives on Medical Education*. 2014;**3**:124-128.

[18]   Bok H, Teunissen P, Spruijt A, et al. Clarifying students' feedback-seeking behaviour in clinical clerkships. *Med Educ*. 2013;**47**:282 - 291.

[19]   Bok HGJ, Jaarsma DADC, Spruijt A, Van Beukelen P, Van Der Vleuten CPM, Teunissen PW. Feedback-giving behaviour in performance evaluations during clinical clerkships. *Medical Teacher*. 2015;**0**:1-8.

[20]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: The potential and limitations of regulatory focus theory. *Medical Education*. 2012;**46**:593-603.

[21]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education*. 2013;**47**:585-594.

[22]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Music lessons: Revealing medicine's learning culture through a comparison with that of music. *Medical Education*. 2013;**47**:842-850.

[23]   Vandenbroucke JP, Elm Ev, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Annals of Internal Medicine*. 2007;**147**:W-163.

[24]   von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Preventive Medicine*. 2007;**45**:247-251.

[25]   Menon S, Winston M, Sullivan G. Workplace-based assessment: survey of psychiatric trainees in Wales. *Psychiatric Bulletin*. 2009;**33**:468-474.

[26]   Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Medical Teacher*. 2011;**33**:919-927.

[27]   Menon S, Winston M, Sullivan G. Workplace based assessment: attitudes and perceptions among consultant trainers and comparison with those of trainees. *The Psychiatrist*. 2012;**36**:16-24.

[28]   Swayamprakasam AP, Segaran A, Allery L. Work-based assessments: making the transition from participation to engagement. *JRSM Open*. 2014;**5**.

[29]   Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*. 2012;**46**:201-215.

[30]   Hassell A, Bullock A, Whitehouse A, Wood L, Jones P, Wall D. Effect of rating scales on scores given to junior doctors in multi-source feedback. *Postgraduate Medical Journal*. 2012;**88**:10-14.

CHAPTER 3

Is the learning value of workplace-based assessment being realised? A qualitative study of trainer and trainee perceptions and experiences

Authors

Aileen Barrett, Rose Galvin, Pim Teunissen, Albert Scherpbier, Ann O'Shaughnessy, Mary Horgan

Abstract

Objectives

Workplace-based assessments (WBA) were originally intended to inform learning and development by structuring effective observation-based feedback. The success of this innovation has not yet been established, due in part to the widely varied tools, implementation strategies and research approaches. Using an emerging conceptual framework of *experience, trajectories and reifications (ETR)* in workplace learning we aimed to explore trainer and trainee experiences and perceptions of the learning value of workplace-based assessments. We also attempted to understand, within the first level of this framework, how their multiple experiences have influenced or shaped their learning trajectories.

Methods

Trainers and trainees who had used at least one WBA in the previous year were invited to participate in semi structured interviews for this phenomenological study. We used a template analysis method to explore and compare the experiences of the two groups, using the emergent themes to develop an understanding of the impact of these experiences on perceptions of learning value.

Results

Nine trainers and eight trainees participated in the study. Common themes emerged among the two groups around issues of responsibility and engagement, along with (mis)understandings of the purpose of the individual tools. Trainer-specific themes emerged related to the concurrent implementation of a new eportfolio and perceptions of increased workload. Trainees associated WBA with a *training* value but in view of their experiences of WBA as a 'tick-box exercise', could not translate experiences into reified *learning* values.

Conclusions

Workplace-based assessment provides trainees with a justified reason to approach trainers for feedback. WBA however, is not being reified as the formative innovation originally intended. A culture change may be required to change the focus of WBA research and reconceptualise this set of tools and methods as a workplace learning *practice.*

Introduction

With the international adoption of outcome-based models of medical education comes an increasing reliance on workplace-based assessment (WBA). It is therefore important to understand what WBA has to offer to learners and teachers in clinical practice. Unfortunately, two decades of research exploring the success of workplace-based assessment tools and practices have resulted in varied and conflicting evidence. It appears that WBA can have some impact on changing 'observable' behaviour, for instance Hiemstra et al[1] demonstrated changes in operative procedure skills using 'learning curves' generated by consecutive OSATS. A change in practice or skills is, however, a relatively simplistic approach to seeking evidence for the complexity of learning development.[2] Whether WBA may also have the potential to positively impact aspects of learning that are more difficult to observe - such as affirmation of good practice and reinforcement of sound critical thinking - has yet to be established. In this study, we aimed to explore perceptions of learning value of WBA from the perspectives of trainers and trainees. Additionally, we set out to understand how their WBA experiences have shaped those perceptions.

Formative assessment and learning

WBA was introduced as a potential solution to the problem of infrequent observation of – and feedback to – medical trainees.[3] The purpose of these low-stakes formative assessments was to provide structure for the delivery of observation-based feedback and to inform trainee learning and development.

The concept of formative assessment - or assessment-*for*-learning - originated in primary and secondary school approaches to teaching and learning.[4] One of the main purposes of formative assessment is to develop self-regulated learners who can independently identify learning needs, formulate a plan to meet those needs and, most significantly, self-monitor the outcome.[5] The critical factor in ensuring an assessment is formative is the provision of effective feedback.[6]

But how these WBAs play out in actual practice remains unclear, partly as a result of the varied ways in which the tools have been implemented. As with many educational innovations, the implementation of WBA has met with some significant challenges and research continues to highlight the consistent 'tick-box' approach to completion, concerns around relevance and responsibility[7-9] and issues of innovation implementation and acceptability.[10, 11]

In attempting to evaluate the impact or effectiveness of formative assessment a number of systematic reviews have concluded that WBA does not affect a demonstrable change in practice.[12, 13] However, by defining learning in this limited, though functional way, other potentially important and less tangible learning effects, e.g. reinforcement of positive behaviours and affirmation of good practice may be undetected. In this study, we attempted to uncover these conceptions in the context of the individual experiences of WBA and whether these experiences have resulted in the realisation of the intended learning value of WBA.

Conceptual framework

The underlying conceptual framework informing our study is Teunissen's '*experience, trajectories and reifications*'(ETR).[14] Developed from research in workplace learning, this framework is described as 'a systems approach to help researchers, educators and clinicians understand some of the features of acting and learning in healthcare workplaces'. The purpose of the framework is to make sense of the way in which an individual's multiple workplace experiences lead them on a particular learning trajectory, which Teunissen defines as "combinations of personal experiences over time that result in personal development and impact the constant negotiation of our identity".[14] Over time, people develop patterns of working with (or around) WBA and in doing so reify – or make a concrete reality of - WBA in actual practice.

In this context, we aimed to explore how trainers' and trainees' experiences have influenced the reification of the learning value of WBA. By attempting to understand both the individual conceptualisations of learning and the impact of

WBA on that learning, we used the basis of experience to allow trainees and trainers to articulate how these perceptions had been formed.

Methods

Study design

We identified two research questions:

1. What are trainers' and trainees' perceptions of the learning value of WBA?
2. How have trainer and trainee WBA experiences shaped these perceptions?

We approached the study from a phenomenological perspective employing an interpretive lens in order to firstly explore the lived experiences of WBA users and then to find out how these experiences had impacted on the participant perceptions. To this end, we conducted a series of guided one-to-one interviews with postgraduate medical trainers and trainees. The COREQ standardised reporting guidelines were followed to conduct and report the research.[15] As the principal investigator in particular is involved in both faculty development and quality improvement initiatives in postgraduate medical education, we acknowledged that our perspectives and positions would influence the study; a phenomenological approach allowed us to recognise and articulate those positions within the research.[16] The principal investigator had recently completed a large-scale audit which found that trainees' engagement with WBA appeared to be limited to completing mandatory requirements. We wanted to find out why this was the case and what, if any learning value was attached to WBAs by those who did (or didn't) engage with WBA.

Setting

The study was carried out among trainers (medical consultants) and trainees registered with one of six postgraduate training bodies housed by the Royal College of Physicians of Ireland in 2013. Workplace-based assessments were incorporated into requirements for all training programmes in 2010 along with a new ePortfolio. The majority of programmes implemented WBAs as formative assessments, but with a minimum number to be completed. One programme used a summative

Direct Observation of Procedural Skills (DOPS) for one component of the procedure-based assessments. The Mini-Clinical Evaluation Exercise (Mini-CEX); DOPS and Case-based Discussion (CbD) were the most commonly implemented WBA tools, along with procedure-specific Objective Structured Assessment of Technical Skills (OSATS) in obstetrics and gynaecology and a number of procedure-specific DOPS assessments in gastroenterology. The WBA is completed on paper and the information transferred to the ePortfolio by the trainee. The system then requests 'sign-off' by the named trainer to ensure accuracy of content. Trainees are required to take responsibility for the initiation of the WBA according to policy guidelines.

Ethics Approval

The study was approved by the Research Ethics Committee of the institution (Appendix 1). Given the small potential pool of participants and the position of the principal investigator (AB) working with trainers and trainees, we included a study gatekeeper for all recruitment processes.

Participants and recruitment

Trainers and trainees who had used at least one WBA in the previous year were invited to participate in interviews exploring their experiences. Recruitment took the form of an invitation email (via a gatekeeper who was not involved in the study) to all registered trainers and trainees along with a research webpage notice and emails to the deans of all six training bodies informing them of the study. We aimed to recruit trainers not only in clinical training roles, but in other institutional roles related to training including National Specialty Directors, Deans and Directors of Education. A recruitment incentive (€20 gift token) was provided to all trainees participating in the study. The study gatekeeper forwarded all expressions of interest in the study to AB who then contacted potential participants to arrange interview dates and times. Recruitment to research studies at our institution has been relatively challenging as our study population is based at clinical sites around the country. We therefore stated in the study information that AB would travel to clinical sites to facilitate participation.

Consent was obtained in two stages; written consent was obtained immediately prior to all interviews. Participants were informed that they would be provided with the opportunity to review their interview transcripts and at that stage, they could also choose to remove their data from the final analysis. Following this stage all participants approved the inclusion of their data in the study.

Interviews

The interviews were conducted by AB and took place in a variety of settings including the training institution or hospitals and one (trainer) interview took place over the phone. The interviews were recorded on a smartphone application and then transferred to a password-protected laptop prior to transcribing and removal of all identifiers.

The interview questions were in part based on previous work examining the role of WBA in postgraduate medical training together with elements that were considered relevant to the study in terms of examining participants' perceptions and experiences of WBA (Appendices 2 & 3). The interviews explored experiences using WBA and therefore were guided, but not structured, allowing the questions to emerge from those experiences. After the first two interviews, AB and RG reviewed the recordings to refine the interviewing technique. At this stage we identified that participants did not have a clear understanding of the purpose of WBA and we agreed to explicitly ask subsequent participants to articulate this understanding or perception.

Analysis

A template analysis method was used to explore the findings. This method facilitates coding of the interview transcripts into hierarchical sets of themes and sub-themes.[17] In this study two distinct groups participated, differentiated by how they use WBA i.e. learners and teachers. Using a template analysis method therefore allowed us to create two individual templates and compare and contrast the main themes emerging from the data.[17] Where the themes converged the template analysis also then allowed us to explore the similarities and differences in the sub-themes between the groups.

AB and RG read all transcripts (Appendix 4; sample transcript) and carried out the initial analysis. As a practitioner-researcher, AB was to some extent 'positioned' within the research while RG was completely removed from the research context and therefore together we brought different perspectives to the analysis process. However, instead of applying a priori codes or sensitising concepts,[18] the coding process took place in 'batches' in which the first four interviews were analysed together to get a sense of the emerging themes. This preliminary coding was then 'clustered' into emerging themes and a draft template was constructed. As subsequent transcripts were analysed, the templates were revised and refined by discussion to construct and agree a hierarchy of major themes and lower-level codes related to the research question and the conceptual framework (Figs 1-4). A memo-writing exercise was also employed throughout the process to allow AB to reflect on any additional issues related to the interview process e.g. participant behaviours.

Results

Participant profile

Nine trainers and eight trainees participated in the study. Two male and seven female trainers participated, along with three female and five male trainees. The trainer profile represented five of the six training bodies and included two National Specialty Directors. All trainees were in their first three years of higher specialist training and had completed their basic specialist training in the institution. Trainees were recruited from four programmes in two training bodies. One trainee was completing a Masters in Clinical Teaching and another had previously completed higher training and worked in another medical specialty. Interviews lasted between 8 and 31 minutes; the shorter interviews took place in a hospital setting where the participants were taking time out of clinical work.

The two templates to emerge from the analysis (Appendices 5 & 6) separated into two distinct sections: WBA experiences and perceptions of learning value. Within these two sections, the templates contained some common themes and others that

were specific to each group. Common themes were those of how WBA was used in real practice and experiences of responsibility and engagement. Differing themes that emerged solely among trainers included the association of WBA with the new eportfolio and additional workload; trainees experienced WBA as a valid way of approaching a trainer for feedback thereby protecting 1:1 'training time'. Perceptions of learning value were also commonly expressed as *potential*, as opposed to realised, values.

The findings are presented here under 'common' and 'differing' themes; trainer and trainee quotations are included to reflect the similarities and differences.

Experiences: Common themes

One of the major themes to emerge from the experiences of both groups was *how WBA was used*. In both groups, the overarching picture to emerge was that of a 'tick-box' exercise, completed towards the end of a year or training post and done to fulfil requirements. Within this theme, another important sub-theme to emerge was an admission – among both groups – that the forms were filled retrospectively and tended to be done on recall of a clinical case or discussion; assessments were not set up prospectively and forms were rarely completed at the time of a patient-trainee interaction or after a case-based discussion:

So they (trainees) would come back at the end of the three months and say 'do you remember Mrs X who was in clinic and you know, who had the family problems and the blah, blah – I've used that as my case discussion (Trainer 2)

These repeated patterns both within and between clinical sites also impacted on perceptions of authenticity of WBA. Trainees felt that a mini-CEX was only related to taking a patient history or physical examination and was something that was 'set up' instead of a real-time observation:

It's (mini-CEX) a formal assessment but the problem becomes I suppose (that) you're trying to manufacture, like, a simulated assessment which ……. I suppose it just feels very artificial (Trainee 4)

A second major theme to emerge from both groups was the concept of *responsibility and engagement*. However, the experiences of these concepts varied both within and between groups. Trainers felt that it was the responsibility of trainees to approach them to complete a WBA and described having to 'chase' trainees to complete requirements:

They're (trainees) not great at making it happen, usually it's me who says 'come on now, we need to do it'. I'm surprised at how, even I have a good trainee at the moment and I'm surprised at how, maybe there's a lot going , it's her exam year and stuff but I'm surprised at, you know, I have to keep bringing it up at our, I'd like her to come in at our weekly meeting…. I'd like her to come in and say 'I'd like to do a CbD for my portfolio'; her portfolio is her responsibility (Trainer 7)

However one trainer felt that it was his role as a trainer to guide trainees through their requirements and help them to map out these requirements, particularly in the earlier stages of training.

Trainees also cited issues of responsibility and engagement and felt that completion of their WBAs should be a joint responsibility. While some trainees had experiences of 'chasing' very busy clinicians, they also noted that a consultant had never refused to carry out an assessment. A number of trainees also cited the nature of the consultant/trainee relationship as either a potential facilitator or barrier to initiating the WBA.

Experiences of feedback were a third major theme to emerge. Trainers generally felt that they provided feedback even in the absence of a WBA while seven of the eight trainees felt that they only experienced feedback on rare occasions and that the value of WBA was in justifying their approaching trainers for feedback:

….for programmes where that isn't occurring, that observed teaching for procedures isn't occurring, where this forces that trainer to have a tool to train the trainee… (Trainee 6)

In order to probe more deeply into feedback experiences, all participants were asked to recall a specific WBA-related instance where they felt the feedback had informed their learning. Only one trainee could recall an experience that informed learning and further development. Another trainee cited an experience of his trainer asking him if he should insert positive feedback into his WBA:

We were filling out the forms at the end of it and my trainer was saying 'gosh, that was very good' and there was a comment box at the end and he said 'I don't really want to put anything in there now because it might be looked on as a criticism'. And I said 'well, you know you can write something positive if you want'! (Trainee 1)

Experiences: Trainer-specific themes

Trainers alone linked WBA experiences to the eportfolio technology and their views of the learning value were linked more closely to costs associated with time and inefficient technology systems. One trainer felt that their role as a trainer was to ensure proficiency in technical skills, but that there wasn't really time to assess communication issues:

In terms of a trainer I would be more focused on teaching specific skills. It's very hard to, I suppose you could use them in terms of communication skills and you know, counselling and that sort of thing. But I really, our specialty isn't great at focusing on that (Trainer 3)

Other trainers felt that the role of the trainer in day-to-day practice negated the need for additional WBA as they were aware of their trainee's progress:

Personally I think it's a bit of a waste of time…. I assess my trainees based on their real-life performance (Trainer 2)

In balancing this view, some trainers felt that while they were busy, that their role as a trainer meant finding the time to complete WBA requirements and to support their trainees.

Experiences: Trainee-specific themes

Trainees articulated a value in WBA in terms of 'forcing' observation and feedback. This theme did not emerge at all within the trainer group. Trainees also felt that WBA represented a form of 'protected training time' and cited issues of balancing service and training that were not articulated by trainers:

I think they're probably a good thing…..that you are receiving training from the college or from your trainer (Trainee 5)

However, one trainer who reported the regular use of WBA also cited issues of a perceived disconnection between concepts of service and learning:

I think they get out of it a specific session with their trainer that leads them to believe their trainer is interested in their training……..I suppose I see the biggest issue is that they don't realise they are being trained all the time during their job. They want specific sessions on training whereas the Mini-CEX, the DOPS, the case-based discussions, this is something they're doing all the time and they don't recognise that as learning (Trainer 6)

Perceptions of learning value

In both groups, perceptions of learning value, while linked to experiences, were mainly discussed in hypothetical terms, particularly among trainees. When asked whether they felt WBA was a valued learning innovation, trainees responded that it had the potential to be valuable if used differently; the *concept* was valuable but had not been realised as such. This potential value was articulated in terms of

diagnostic assessment, affirming good practice (as well as required improvements) and also in anchoring performance. However, the perceptions of the barriers to the realisation of this potential value appeared to be also tied into the issues of responsibility and engagement, along with trainee perceptions that the trainers often required 'guidance' in how to use the tools and that they themselves were unclear as to how to use these tools.

A number of trainers and trainees who had had experiences in other training jurisdictions suggested that 360⁰/multisource feedback could provide very valuable learning information; this tool was perceived by both groups to provide a mechanism for delivering feedback on overall performance as a doctor that was not threatening and more inclusive of all the doctors' roles within a multidisciplinary team:

… you wouldn't want it too formal but I think it'd be no harm, let's say, if your trainer's getting feedback from other professionals in terms of nursing, physio, OT etc., how they're finding your interactions, you know? Because that's such a crucial element of how you're going to function as a consultant (Trainee 1)

Discussion

The experiences of workplace-based assessment (WBA) among this group of trainers and trainees in many ways mirrors those of studies in other medical disciplines and training jurisdictions. Where WBA was perceived to be an academy-driven, 'sign-off' initiative, trainers' and trainees' attachment of learning value to WBA was limited. However, a number of additional themes emerged.

Given the 'bad press' WBA has had in the recent literature, particularly in studies related to trainee attitudes and perceptions,[19-22] trainees in our study still viewed WBA as having potential learning value. However, in analysing these positive perceptions, it appeared that this value was actually articulated as *training value* i.e. facilitating time with their trainer, and viewed as protected teaching time. These issues have begun to emerge in recent literature and have their roots in tensions between the perceptions of trainees of their role and/or identities as trainees and service providers. Traditional models of postgraduate education in medical specialties have always relied on the provision of service as the basis for training in an apprenticeship-style programme and according to Kesselheim and Cassel,[23] service and education 'have been set in conflict' since the formalisation of postgraduate education. With the increasing demands placed on trainees as service providers within healthcare systems, internationally the tension between these two constructs also appears to be increasing.[24, 25] In our study trainees viewed trainers' engagement and willingness to participate in WBA as an indicator of training input; however they found it difficult to articulate or describe experiences of a specific learning value.

Trainer experiences in general were dominated by issues of trainee engagement and responsibility and when asked, trainers could not link their experiences of delivering WBA feedback to a concrete example of learning. This may be explained by the complexities of defining 'learning'.[2] Teunissen[14] also suggests that learning can also result from affirmation of practice and the need to change practice in the context of positive feedback may therefore be a limited marker of 'learning'.

This study set out to not only ascertain perceptions of trainers and trainees as to the learning value of WBA, but to also attempt to understand how their multiple experiences, in various situations, have informed those perceptions. In exploring this link, we used Teunissen's *Experience-Trajectories-Reifications* conceptual framework of workplace learning[14] as a basis for understanding the links between implementation and experiences. However, throughout this study, as with others, [8, 10] it became clear that WBA has not yet become part of the fabric of workplaces as learning environments as was intended and it is as yet difficult to evaluate the impact of WBA on the trajectories of trainers and trainees. As a result, we found that our participants mainly addressed the first level of the ETR framework.

In this study, both trainees and trainers had had limited experience of WBA as assessments tended to only be completed towards the end of the academic year or clinical post. The varying and limited number of WBA experiences limits our ability to interpret and plot (or track) trajectories arising from these experiences; we could therefore only evaluate the findings in the context of the first level of the ETR framework. While this limitation may be related to the specific context in which the WBA has been implemented it highlights the need for more longitudinal studies to explore the educational impact of WBA.

The issues surrounding authenticity that emerged from the data may also have impacted on the perceptions of learning value. This particular finding was surprising – the basis of WBA is that is situated in an authentic context and therefore is different from a simulated setting, for example, as with the OSCE format. In 2003, when the OSCE was riding a wave of popularity in undergraduate medical education, Bryan Hodges[26] queried whether or not the OSCE format created an 'authentic situation' and used Goffman's theory of 'the presentation of the self'[27] to propose that in this assessment format, the learner chooses how he/she presents him/herself much like an actor in a play. In Goffman's theory, people have a 'back stage' and 'front stage' persona and continually make choices as to which persona the public will get to see. Translating this to medical education assessment practices, Hodges suggested that in an OSCE setting, the learner chooses whether

to perform for the patient as they would in a real-life setting or whether to perform for the assessors and their perceptions of the expectations of the assessors.

In our study, the issues surrounding the way in which WBA is implemented – including the 'staging' of a WBA event - may have impacted on how the trainee chose to present themselves. While we didn't explore this finding fully in this study, it is another factor worth considering and whether this may impact in general on the perceptions of WBA among its users or if this a local implementation factor.

Given these experiences and trajectories, it appears that the reification of learning value in WBA has not happened in a way that is in line with its espoused purpose. The reification of WBA as a 'tick-box' to fulfil requirements has instead negatively impacted on the relationship between learning and WBA. However, the experiences have not completely negated the potential for this learning value to be realised; rather changes to the situations in which assessments are completed may impact more positively on this reification in the future. Trainers and trainees did not identify any inherent issues with the design of the tools, rather the way in which they are used. A recent UK study [28]on the implementation of Supervised Learning Events (SLEs) raised a number of similar issues of confusion around the purpose of the innovation as opposed to the tools involved. Interestingly the tools used for the SLEs were the same workplace-based assessment tools used previously, but re-named.

An important issue to emerge from this study and worthy of future study includes approaching this topic from more narrative perspectives, including that of critical discourse analysis. In this study, issues of responsibility to change the learning value from potential to realised were raised, however we did not fully explore why participants felt that they could (or could not) influence this change.

These findings have important implications for practice in postgraduate medical education. While some of the findings relate to the specific implementation issues locally, they also highlight another issue worthy of consideration – that of

reconceptualising workplace-based assessment as a *practice*, supported by well-designed tools, rather than a set of tools hindered by poorly designed practices. We have shown that trainees continue to seek feedback on their performance – both for affirmation and constructively critical purposes. We have shown that there is a role for WBA in providing an avenue to this feedback based on direct observation. We hope to now continue the international conversation on the place and position of WBA in programmes of assessment for postgraduate training by reframing the practice of workplace-based assessment as the effective implementation of direct observation of practice and related feedback.

Limitations of the study

This study was initiated with the aim of adding to the body of knowledge on the implementation of WBA internationally. While the findings were thoroughly and rigorously analysed, this is a phenomenological study and we therefore recognise that we did not set out to achieve the 'truth' of WBA experiences and implementation. Instead, we wished to use our study to inform quality improvement initiatives and practices within our own and other institutions by understanding the specific as well as the international experiences and contexts. We also recognise that in recruiting to an interview-based study respondents generally may have been more engaged – albeit positively or negatively – in the debate on WBA than non-responders. We endeavoured to also ensure a representative spread of disciplines and potential perspectives but the sample remained small. One of the main learning points regarding recruitment from this study has been to consider the use of social media to engage future potential study participants.

Conclusion

The implementation of workplace-based assessment tools and methods has proved challenging. As educational research methods and practices evolve, it appears that a purely psychometric approach to evaluating their purpose as either formative *or* summative tools may be insufficient in evaluating their impact. In this study, trainees continue to value workplace-based assessment; however where the intention is to identify and create opportunities for continuous assessment, feedback and deliberate practice informed by good feedback, perhaps the concept of formative assessment may be better conceptualised as a *practice*.

Authors' Contributions

AB conceived the study and the original research questions, carried out the interviews, coordinated the data analysis, drafted the paper and coordinated edits. RG collaborated on all stages of the study including the design of the study, the interview schedule, data analysis and the paper write-up. AJJS collaborated on all aspects of the study design and analysis and so-supervised the PI. PWT advised on the conceptual framework and contributed to the final analysis and write-up. AO'S collaborated on the study design, ethics application and final write-up. MH advised on all aspects of the study and supervised the PI. All authors approved the final draft.

References

[1]   Hiemstra E, Kolkman W, Wolterbeek R, Trimbos B, Jansen FW. Value of an objective assessment tool in the operating room. *Canadian Journal of Surgery*. 2011;**54**:116-122.

[2]   De Houwer J, Barnes-Holmes D, Moors A. What is learning? On the nature and merits of a functional definition of learning. *Psychon Bull Rev*. 2013;**20**:631-642.

[3]   Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (Clinical Evaluation Exercise): A Preliminary Investigation. *Annals of Internal Medicine*. 1995;**123**:795-799.

[4]   Scaife J, Wellington J. Varying perspectives and practices in formative and diagnostic assessment: a case study. *Journal of Education for Teaching*. 2010;**36**:137-151.

[5]   Zimmerman BJ. Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*. 2000;**25**:82-91.

[6]   Clark I. Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*. 2012:1-45.

[7]   Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Medical Teacher*. 2011;**33**:919-927.

[8]   Bok H, Teunissen P, Favier R, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education*. 2013;**13**:123.

[9]   Cohen SN, Farrant PB, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol*. 2009;**161**:34-39.

[10]   Fokkema JPI, Teunissen PW, Westerman M, et al. Exploration of perceived effects of innovations in postgraduate medical education. *Medical Education*. 2013;**47**:271-281.

[11]   Fokkema JP, Scheele F, Westerman M, et al. Perceived effects of innovations in postgraduate medical education: a Q study focusing on workplace-based assessment. *Academic Medicine*. 2014;**89**:1259-1266.

[12]   Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;**341**:c5064.

[13]   Overeem K, Faber MJ, Arah OA, et al. Doctor performance assessment in daily practice: does it help doctors or not? A systematic review. *Medical Education*. 2007;**41**:1039-1049.

[14]   Teunissen P. Experience, trajectories, and reifications: an emerging framework of practice-based learning in healthcare workplaces. *Advances in Health Sciences Education*. 2014:1-14.

[15]   Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care* 2007;**19**:349-357.

[16]   McLachlan E, King N, Wenger E, Dornan T. Phenomenological analysis of patient experiences of medical student teaching encounters. *Medical Education*. 2012;**46**:963-973.

[17]   King N. Doing template analysis. In: Symon G, Cassell C, eds. *Qualitiative Organisational Research: Core Methods and Current Challenges*. AGE, London, UK, 2012; 426-450.

[18]   King N, Horrocks C. *Interviews in Qualitative Research*. London, UK SAGE Publications Ltd; 2010.

[19]   Menon S, Winston M, Sullivan G. Workplace-based assessment: survey of psychiatric trainees in Wales. *Psychiatric Bulletin*. 2009;**33**:468-474.

[20]   Menon S, Winston M, Sullivan G. Workplace based assessment: attitudes and perceptions among consultant trainers and comparison with those of trainees. *The Psychiatrist*. 2012;**36**:16-24.

[21]   Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Medical teacher*. 2011;**33**:919-927.

[22]   Tailor A, Dubrey S, Das S. Opinions of the ePortfolio and workplace-based assessments: a survey of core medical trainees and their supervisors. *Clinical Medicine*. 2014;**14**:510-516.

[23]   Kesselheim JC, Cassel CK. Service: An Essential Component of Graduate Medical Education. *New England Journal of Medicine*. 2013;**368**:500-501.

[24]   Sanfey H, Cofer J, Hiatt JR, et al. Service or education: In the eye of the beholder. *Archives of Surgery*. 2011;**146**:1389-1395.

[25]   MacCraith B. Strategic review of medical training and career structure; Report on medical career structures and pathways following completion of specialist training Department of Health, 2014.

[26]   Hodges B. OSCE! Variations on a theme by Harden. *Medical Education*. 2003;**37**:1134-1140.

[27]   Goffman E. *The Presentation of Self in Everyday Life*. London, New York Double Day; 1959.

[28]   Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014;**4**.

CHAPTER 4

Protocol: A BEME review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees

Authors: Aileen Barrett, Rose Galvin, Yvonne Steinert, Ann O'Shaughnessy, Albert Scherpbier, Mary Horgan, Tanya Horsley

Abstract

Background

Workplace-based assessment was designed to facilitate observation and to structure feedback on the performance of trainees in real-time clinical settings and scenarios. Research in workplace-based assessment has primarily centred on understanding psychometric qualities and performance improvement impacts of trainees generally.

An area that is far less understood is the use of workplace-based assessment for trainees who may not be performing at expected or desired standards, referred to within the literature as trainees 'in difficulty' or 'underperforming'. In healthcare systems that increasingly depend on service provided by junior doctors, early detection (and remediation) of poor performance is essential. However, barriers to successful implementation of workplace-based assessment (WBA) in this context include a misunderstanding of the use and purpose of these formative assessment tools.

This review aims to explore the impact - or effectiveness - of workplace-based assessment on the identification of poor performance and to determine those conditions that support and enable detection, *i.e.* whether by routine or targeted

use where poor performance is suspected. The review also aims to explore what effect (if any) the use of WBA may have on remediation or on changing clinical practice. The personal impact of the detection of poor performance on trainees and/or trainers may also be explored.

Methods/design

Using BEME (Best Evidence in Medical Education) Collaboration review guidelines, eight databases will be searched for English-language records. Studies examining interventions for workplace-based assessment either routinely or in relation to poor performance will be included. Independent agreement (kappa .80) will be achieved using a randomly selected set of records prior to commencement of screening and data extraction using a BEME coding sheet modified as applicable[1] as this has been used in previous WBA systematic reviews[2] allowing for more rigorous comparisons with the published literature. Educational outcomes will be organised using Kirkpatrick's framework of educational outcomes incorporating adaptations by Barr[3] and Steinert et al[4] for medical education research.

Discussion

Our study will contribute to an ongoing international debate regarding the applicability of workplace-based assessments as a meaningful formative assessment approach within the context of postgraduate medical education.

Systematic review registration

The review has been registered by the BEME Collaboration www.bemecollaboration.org (Appendix 1)

Keywords

Workplace-based assessment; formative assessment; postgraduate medical education; residency training; poor performance; remediation; systematic review

Background to the topic

Poor performance among junior doctors has a potentially significant impact on patient safety and care. In an extremely pressured healthcare system that depends to a very large extent on the service provided by junior doctors, early detection (and remediation) of poor performance is essential. Studies carried out in the late 1980s and early 1990s articulated, for the first time, that doctors-in-training were very rarely provided with feedback and even more rarely observed in practice.[5, 6] In 1995 the first workplace-based assessment tool specifically designed to structure feedback following an observation of a clinical interaction – the Mini-Clinical Evaluation Exercise (Mini-CEX) was developed[7] using emerging theory from the UK and elsewhere on assessment-*for*-learning in which the goal of the interaction is to provide feedback on performance and inform a learning plan or action, and which may or may not involve the award of a grade or mark.

Many more tools have since been developed and while workplace-based assessment (WBA) was originally mooted as a formative assessment strategy, it has been adopted primarily in many countries as a quality assurance mechanism, detailing and tracking progress in order to inform end-of-year sign-off. As such, most of the research has focused on whether or not the tools used in WBA are valid and reliable. These primary research studies have either reviewed the reliability and validity of existing tools in different areas of clinical practice e.g. emergency departments or validated development of new tools designed to evaluate specific aspects of practice or performance.[8, 9] As of 2012, more than 52 WBA tools had been designed and published.[10]

An emerging area of investigation now relates to profile issues with feedback and why its impact may be limited and how trainees perceive that feedback.[11, 12] Literature and anecdotal evidence suggests that trainers feel uncomfortable giving negative feedback and structuring learning plans for trainees[11] and that trainees view WBA as a 'tick-box exercise'[12] with little impact on their learning and development. More recently, Watling et al[13] investigated the complexity of the impact of feedback by exploring how trainees process the information and decide

what to do with it. This paper highlights the role of 'credibility judgements'; trainees make a judgement on the clinical expertise of the person providing the feedback to decide whether or not the feedback is relevant and if they should accept it.

This may lead to problems where a trainee is underperforming and a trainer is struggling to identify the issues at the source of the problem and to establish whether these are issues of learning or relate to health and other concerns. Black and Welch [14] reported that of 60 doctors identified as underperforming (in a deanery of 1482 Foundation Year 1 and 2 trainees), 16.6% of them were identified using a mini-PAT (mini Peer Assessment Tool) workplace-based assessment alone, while the remainder were identified by trainer observation of performance and reporting of health-related issues. In this case, formalised workplace-based assessments were no more effective than trainer observation; it remains unclear as to whether these underperforming trainees would have been identified without any formalised WBA process.

In the UK, the WBA process is trainee-led therefore trainees choose their assessor and request a WBA. A recent UK-based study also explored whether trainees 'in difficulty' use WBA differently to their peers.[15] The researchers did not find strong associations between trainees in difficulty and the level of complexity of the clinical cases they used for a WBA but they reported strong associations between those trainees and the assessors they chose to carry out the WBA. Furthermore, trainees in difficulty were more likely to approach a nursing colleague to complete a DOPS (direct observation of procedural skills) assessment and a non-clinical assessor to carry out a mini-PAT (mini-peer assessment tool) possibly indicating some level of avoidance of medical peers and senior colleagues among those trainees with insight into the fact that they were underperforming. However, whether or not they approached these assessors after they had been deemed to be 'in difficulty' is not clear.

There have been a number of published systematic reviews in the area of workplace-based assessment and its impact, either on learning or performance.

While the studies all varied in the focus of the research question, all of the following reviews cited the difficulty in determining any conclusive findings as the published literature to date varies so significantly in terms of methods and quality.

- Kogan et al[10] reviewed the psychometric properties and validity of 52 existing workplace-based assessment tools. One of the problems they identified in carrying out the review was the lack of methodological homogeneity, making comparisons between tools - and the educational effectiveness of individual tools - too complex to evaluate

- Miller and Archer[2] explored the impact of Mini-CEX, DOPS, case-based discussion and 360° multisource feedback on performance and concluded that peer assessment (360°) had some impact on changing practice but similarly, that differing study methodologies and reported outcomes limited the generalizability of individual study results

- Overeem et al[16] performed a review of 64 articles detailing methods of performance assessment in the clinical setting. They determined that while a number of methods may be feasible (particularly peer assessment), the effectiveness of formative assessments in influencing changes in performance is limited

- Saedon et al[17] reviewed 15 prospective studies of workplace-based assessments in postgraduate medical education. They too were unable to ascertain a definite link between workplace-based assessment and an improvement or change in performance. An important aspect of this review, according to the authors, was that the majority of the studies included used self-reported - as opposed to externally observed – changes in practice

- A literature review by Pelgrim et al,[18] exploring the reliability and validity of single-patient encounter observation tools concluded that while the

instruments reviewed appeared to demonstrate a good level of feasibility, the Mini-Clinical Evaluation Exercise was the only tool with enough evidence to demonstrate 'acceptable' reliability, and this was over ten encounters. While this study did not explore changes in performance, it raises the question of the use of formative assessments as assessments of performance and in high stakes situations (e.g in making progress judgments), these tools have yet to be fully validated.

However, these systematic reviews did not review studies looking solely at the changes from a baseline of poor or underperformance. One of the criticisms of WBA, and a feature we have recently identified[19] is the potential 'ceiling effect' of WBA rating systems – if a competence or aspect of performance is deemed to be 'meeting' or 'above expectation', a change in practice may be less likely to occur. Our review therefore aims to progress this work and clarify how WBA might affect performance specifically in underperforming or poorly peforming trainees.

An initial literature search has identified a number of studies looking at the identification of poor performance using specific tools;[14, 15] however we are not yet aware of any systematic review that has explored the use of WBA in general as a method of identifying or remediating poor performance among postgraduate medical trainees.

The aim of this review therefore, is to explore whether or not workplace-based assessment can assist in identifying or remediating poor performance, and the conditions under which WBA may be of use in this purpose.

Review Question, Objectives and Key Words

Primary Review Question

Can workplace-based assessment be used to identify and remediate poor performance among postgraduate medical trainees?

Secondary Question

What features of workplace-based assessment tools and/or factors associated with WBA methods and utilisation primarily contribute to the usefulness of WBA in identifying or remediating poor performance among postgraduate medical trainees?

Definitions:

- **Underperformance** within a clinical context is inconsistent within the literature and terms are often used interchangeably. The most contemporary (2013) definition provided within a UK-based study defines the underperforming trainee as 'requiring intervention beyond the normal level of supervisor-trainee interaction'.[15] While this is the most recent definition, it does not classify the root cause of the trainee's difficulties; rather it provides an overarching articulation of a trainee who is not currently meeting the expectations of their training level. Building on this definition, and in an attempt to ensure comprehensive inclusion, we will also include derivations of the concept that include:

  - The trainee in difficulty
  - The difficult trainee
  - The trainee in trouble

- For the purposes of our review **workplace-based assessment** will be defined as any assessment tool or method designed to provide feedback on performance and improvement practice in a clinical setting

- **Postgraduate medical trainees** are post-registration doctors enrolled in a training programme e.g. in medicine or surgery (e.g. resident, trainee, doctor-in-training, non-consultant hospital doctor)

- **Remediation** is 'the act or process of correcting a deficiency' (p e185) as described by Cleland et al (2010).[20] This particular definition was chosen for our review as it links closely with the purpose of formative assessment, which is to provide information on performance strengths and deficiencies, and to provide a structure for feedback and guidance on improving performance

- **Features** or **factors** of workplace-based assessment:
    - WBA assessment tools will include rating systems and feedback structures
    - WBA methods of use include such considerations as whether they are used routinely or in the case of suspected underperformance, if multiple tools or single ones are used and if one or many encounters were used in identifying or remediating performance-related issues

Objectives

Primary objectives

1. To determine if the workplace based assessment can be used to identify and/or remediate underperformance among postgraduate medical trainees

Secondary objectives

1. To determine the conditions of use under which WBA may assist in identifying or remediating poor performance
2. To identify features of WBA tools, or factors in using WBA, that are most likely to contribute to successful remediation of poor performance

Educational outcomes will be organised using Kirkpatrick's framework of educational outcomes using Barr's adaptations[3] for medical education research

(Table 1). We will also use the modification of level 3 (change in behaviour) proposed by Steinert et al[4] to distinguish between self-reported and observed changes.

Table 1: Kirkpatrick's framework of educational outcomes

Level I	Learner's reactions
Level 2a	Modification of attitudes and skills
Level 2b	Acquisition of knowledge and skills
Level 3a	Self-reported change in behaviour
Level 3b	Observed change in behaviour
Level 4a	Change in organisational practice
Level 4b	Benefits to clients or patients

We propose to use these levels although recent criticism of this framework has proposed that it uses a hierarchical approach to evaluating evidence[21] and implies that a change in patient care is the ultimate 'evidence' of learning impact, whereas the importance of a change in learner behaviour at Level 3 may be as important as demonstrating a change in patient care. In this review, we are aiming to ascertain the usefulness of a particular method of learning development in changing those individual-level – as well as system-level - outcomes therefore we will treat the framework as a classification system as opposed to a hierarchical framework of evidence.

Search Sources and Strategies

Working with a Master's of Library and Information Science (MLIS) librarian, those experienced with searching, content experts and published evidence, we will iteratively develop search strings using MESH (medical subject headings) and free-text terms to ensure breadth and depth of coverage (Sample MEDLINE search, Appendix 2). Once the search has been tested and validated, all electronic databases (see below) will be searched to identify potentially relevant records using appropriate derivatives of the searches. Prior to final searching, the foundational search in MEDLINE will be peer reviewed by a PhD-level information scientist using the PRESS: Peer Review of Search Strategies model.

To ensure comprehensiveness of our search the following electronic databases will be searched:

- Medline
- CINAHL
- British Education Index
- EMBASE
- ERIC
- Australian Education Index
- PsycINFO
- Science Direct

Our searches will be limited to 1995 to the most recent search date given thefirst workplace-based assessment study was published in 1995. No limits for study design or publication type will apply. Only English-, French-, German- and Dutch-language reports will be considered for inclusion and were chosen to reflect the abilities of the review authors.

Using a pre-defined concept mapping framework and keywords defined by database thesauri, the following will be considered:

- Postgraduate medical education and all derivatives identified (e.g. postgraduate medical education; residency training)
- Workplace-based assessment(s) and all derivatives identified (e.g. formative assessment, assessment-for-learning)
- Physicians in training, trainee doctors, doctors in training, junior doctors
- Direct Observation of Procedural Skills [DOPS]
- Mini-Clinical Evaluation Exercise, Mini-CEX
- Case-based Discussion, CbD
- Observed Structured Assessment of Technical Skills, OSATS
- Mini Peer Assessment Tool, Mini-PAT
- 360° multisource feedback, MSF
- Formative assessment, assessment-for-learning
- Poorly performing trainee; underperforming trainee; trainees in difficulty; doctors in difficulty
- Performance deficit, academic difficulty

Database searches will be supplemented to ensure comprehensiveness and will include:

1. Additional searches of BEME published reviews, Cochrane database and DARE
2. Review of reference lists of included studies and review articles (http://www.ncbi.nlm.nih.gov/pubmed/21833989)
3. Contact with prominent authors in the field of workplace-based assessment for expert recommendations and guidance and to identify unpublished (including doctoral theses), recently published or ongoing studies relevant to this review. Prominent authors are defined as those having published previous relevant reviews multiple WBA studies

4. Conference presentations from the Association of Medical Education in Europe, Association for the Study of Medical Education, International Conference on Residency Education and Canadian Conference on Medical Education will be searched for relevant abstracts from 1995 forward or the inaugural year of the conference as applicable.

5. We will also conduct a citation search on Web of Science looking for studies citing any of the included articles

Study Selection Criteria

Types of studies to be considered

No restrictions for study design will be applied; qualitative and quantitative studies will be included. However, non-research publications including commentaries, letters and editorials will not be included in the review.

Inclusion/Exclusion Criteria

The inclusion and exclusion criteria for the review are summarised in Table 2.

Table 2: Inclusion/exclusion criteria

	INCLUSION CRITERIA	EXCLUSION CRITERIA
Population	• Postgraduate medical trainees • Postgraduate surgical trainees	• Non-medical trainees • Medical students (undergraduate and graduate-entry programmes) • Studies not involving humans • Clinical studies • Studies not involving physicians
Intervention	• Workplace-based assessment e.g. Mini-CEX, DOPS in direct observation situations (see Types of Interventions)	
Outcomes	• Studies that describe/report outcomes related to identification/remediation (see Types of Outcomes)	
Research Design	Studies which provide primary data for any of the outcomes above including, but not limited to, the following designs: • Experimental and/or observational studies • Randomised and non-randomised studies • Prospective or retrospective cohort studies • Qualitative • Descriptive	• Studies that do not report primary data including commentaries, letters and editorials • Reports published only in dissertation and abstract format

Types of interventions

The interventions which will be considered for this review are those which involve the use of workplace-based assessment either routinely (e.g. as a component of clinical rotations), or in relation to poor performance (e.g. confirmation of poor performance)

We will include studies that describe or evaluate the use of WBA within the context of:

- Routine or targeted use of WBA

- Trainee-led or trainer-led WBA

- Single or multiple-use of WBA tools

- Management or remediation of poor performance for knowledge, skills and attitudes

- Use of WBA as part of a wider programme of assessment in the context of a range of assessment evidence

- Presence/absence of facilitation and/or written or verbal feedback

Outcomes

At least one of the following outcomes must be reported as being specifically resultant from engagement in a component part of a WBA assessment process:

Individual-level outcomes

- Number of trainees identified as poorly performing through the use (either routine or targeted) of a WBA process

- Progression/remediation statistics

- Changes in trainee performance (knowledge, skills, attitudes etc.)

- Trainee satisfaction

Practice-level outcomes

- Changes in implementation methods e.g. non-routine to routine
- Implementation of new/differing WBA tools

System-level outcomes

- Changes in system-wide implementation of WBA tools or methods e.g. throughout a deanery

Secondary outcomes will include the conditions under which the use of WBA is most useful in identifying or remediating underperformance and, where possible, the features of WBA tools, or factors in using WBA, that are most likely to contribute to successful remediation of underperformance.

Procedure for extracting data

Two study authors (AB and RG) will independently review 20% of the retrieved articles (randomly selected) each using a modified BEME Coding Sheet developed a priori to ensure comprehensiveness of the tool. The reviewers' data extraction will be validated for accuracy by at least one other author (TH) for inter-rater reliability to a kappa of .80 agreement. Once this agreement has been reached, all extractions will be completed in duplicate. Discordance with extractions will be resolved through discussion and resolved using a third author when applicable.

The following data will be extracted independently by two authors (AB and RG) and entered into the validated data extraction form:

- Publication characteristics (publication type, journal of publication, year etc.)
- Inclusion of a conceptual or theoretical framework to develop the study
- Study-level Characteristics
    - Population
    - Study design
    - Analysis framework
    - Notable inclusion / exclusion
    - Country/jurisdiction
    - Context (e.g. tertiary-care centre, in situ simulation
    - Unit of analysis
    - Mode of recruitment
    - Sample size
    - Data collection methods
- 'Intervention' details
    - WBA tool(s) and/or description
        - Name (of tool)
        - Content
        - Modifications
        - Frequency of use
        - Routine or targeted use of WBA
        - Trainee-led or trainer-led WBA
        - Single or multiple-use of WBA tools
    - For 'poor performance' or 'remediation', are reliability/validity data stated
- Descriptive / Contextual data
- Risk of bias assessment

- o Have the authors carried out a risk of bias assessment in declaring, for example, author positionality (is the researcher involved in the implementation/use of WBA?)
- Outcome data
  - o Trainee-level
    - Changes in knowledge, skills and attitudes
    - Progression/non-progression to next training stage/rotation/post
    - Remediation outcomes including delayed progression, repeated training stage, evidence of improved performance and reassessment
  - o Practice-level
    - Changes from routine to targeted (and vice versa) use of WBA
  - o System-level
    - System-wide outcomes including changes to deanery/college-wide practice in WBA implementation e.g.
      - Modification of WBA tools
      - Implementation/introduction of new WBA tools
      - Elimination of WBA tools from use throughout the deanery

To ensure security of copyrighted and proprietary materials given the geographically disperse team, EndNoteWeb library will be used to protect and share records fluidly.

Methodological Quality

Methodological quality will be evaluated using the BEME criteria[1] as this has been used in previous WBA systematic reviews[2] allowing for more rigorous comparisons with the published literature. Recognising limitations around reporting quality, we will include a formal risk of bias assessment; we propose to modify one of the BEME quality criteria ('control for confounding') to include author 'positionality' and risk of bias assessment (Table 3).

Table 3: BEME Quality Indicators

Research Question	Is the research question or hypothesis clearly stated?
Study Subjects	Is the subject group appropriate for the study being carried out?
Data Collection Methods	Are the methods used appropriate for the research question and context?
Completeness of data	Attrition rates/acceptable questionnaire response rates
Risk of bias assessment	Is a statement of author positionality and a risk of bias assessment included?
Analysis of results	Are the statistical and other methods of results analysis used appropriate?
Conclusions	Is it clear that the data justify the conclusions drawn?
Reproducibility	Could the study be repeated by other researchers?
Prospective	Is the study prospective?
Ethical Issues	Are all ethical issues articulated and managed appropriately?
Triangulation	Were results supported by data from more than one source?

Synthesis of Extracted Evidence

The study data will be analysed and classified according to the primary and secondary outcomes identified.

Based on our literature search to date and the consistent conclusions of the systematic reviews discussed earlier, one of the most significant challenges in appraising WBA literature is the lack of homogeneity between study methods. We anticipate that heterogeneity may be present within our sub-set of literature and thus meta-analysis is unlikely. However, the team plans to explore and quantify heterogeneity using standard test of heterogeneity (e.g. I^2) and visually using funnel plots to identify and explore outliers. Descriptive synthesis, as described by Saedon et al[17] will also be considered. In the event that heterogeneity of studies precludes quantitative syntheses (*e.g.* extensive subject or statistical heterogeneity), a rich descriptive synthesis including post hoc, exploratory work that attempts to explain differences in findings[22] will be undertaken.

In the case of qualitative studies included for analysis, we will use a qualitative meta-synthesis analysis method[23] to explore the common themes and concepts emerging from the research studies.

Authors' contributions

AB conceived the review and the original research questions, drafted the protocol and coordinated the team reviews of each draft. RG collaborated on the design of the review, reviewed and edited the protocol, collaborated on the design of the search strategy and in the selection of the methods. YS reviewed and edited the protocol, advised on BEME guidelines and structure and advised on the selection of methodological quality assessment criteria. AS reviewed and edited the protocol and advised on BEME guidelines and structure. AO'S reviewed and edited the protocol. MH reviewed and edited the protocol. TH advised on all aspects of design, development and writing.

References

[1]   Buckley S, Coleman J, Davison I, et al. The educational effects of portfolios on undergraduate student learning: A Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. *Medical Teacher*. 2009;**31**:282-298.

[2]   Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;**341**.

[3]   Barr H, Freeth D, Hammick M, Koppel I, Reeves S. *Evaluations of Interprofessional Education: A United Kingdom Review of Health and Social Care.* London CAIPE; 2000.

[4]   Steinert Y, Mann K, Centeno A, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher*. 2006;**28**:497-526.

[5]   Daelmans HEM, Hoogenboom RJI, Donker AJM, Scherpbier AJJA, Stehouwer CDA, van der Vleuten CPM. Effectiveness of clinical rotations as a learning environment for achieving competences. *Medical Teacher*. 2004;**26**:305-312.

[6]   Day S, Grosso L, Norcini J, Blank L, Swanson D, Horne M. Residents' perception of evaluation procedures used by their training program. *Journal of General Internal Medicine*. 1990;**5**:421-426.

[7]   Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (Clinical Evaluation Exercise): A Preliminary Investigation. *Annals of Internal Medicine*. 1995;**123**:795-799.

[8]   Cruess R, McIlroy JH, Cruess S, Ginsburg S, Steinert Y. The Professionalism Mini-Evaluation Exercise: A Preliminary Investigation. *Academic Medicine*. 2006;**81**:S74-S78.

[9]   Young JQ, Lieu S, O'Sullivan P, Tong L. Development and initial testing of a structured clinical observation tool to assess pharmacotherapy competence. *Academic Psychiatry*. 2011;**35**:27-34.

[10]   Kogan JR, Holmboe ES, Hauer KE. Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees. *JAMA: The Journal of the American Medical Association*. 2009;**302**:1316-1326.

[11]   Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*. 2012;**46**:201-215.

[12]   Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Medical Teacher*. 2011;**33**:919-927.

[13]   Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgements. *Medical Education*. 2012;**46**:192-200.

[14]   Black D, Welch J. The under-performing trainee – concerns and challenges for medical educators. *The Clinical Teacher*. 2009;**6**:79-82.

[15]   Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments in Foundation Programme training: do trainees in difficulty use them differently? *Medical Education*. 2013;**47**:292-300.

[16]   Overeem K, Wollersheim H, Driessen E, et al. Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study. *Medical Education*. 2009;**43**:874-882.

[17]   Saedon H, Salleh S, Balakrishnan A, Imray C, Saedon M. The role of feedback in improving the effectiveness of workplace based assessments: a systematic review. *BMC Medical Education*. 2012;**12**:25.

[18]   Pelgrim EAM, Kramer AWM, Mokkink HGA, den Elsen L, Grol RPTM, Vleuten CPM. In-training assessment using direct observation of single-patient encounters: a literature review. *Advances in Health Sciences Education*. 2011;**16**:131-142.

[19]   Barrett A, Galvin R, Steinert Y, et al. Profiling postgraduate workplace-based assessment implementation in Ireland: a retrospective cohort study. *SpringerPlus*. 2016;**5**:1-7.

[20]   Cleland J, Mackenzie RK, Ross S, Sinclair HK, Lee AJ. A remedial intervention linked to a formative assessment is effective in terms of improving student performance in subsequent degree examinations. *Medical Teacher*. 2010;**32**:e185-e190.

[21]   Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence'. *Medical Education*. 2012;**46**:97-106.

[22]   Cook DA, West CP. Conducting systematic reviews in medical education: a stepwise approach. *Medical Education*. 2012;**46**:943-952.

[23]   Walsh ME, Galvin R, Loughnane C, Macey C, Horgan NF. Factors associated with community reintegration in the first year after stroke: a qualitative meta-synthesis. *Disability and Rehabilitation*. 2014;**0**:1-10.

CHAPTER 5

A BEME (Best Evidence in Medical Education) review of the use of
workplace-based assessment in identifying and remediating poor
performance among postgraduate medical trainees

Authors

Aileen Barrett, Rose Galvin, Yvonne Steinert, Albert Scherpbier, Ann O'Shaughnessy,
Mary Horgan, Tanya Horsley

Abstract

Introduction

The use of workplace-based assessment as a facilitator of change in practice among
doctors has not been established; this is particularly important in the case of
underperforming trainees. The aim of this review is to examine the use of WBA in
identifying and remediating practice among this cohort.

Methods

Following publication of a study protocol a comprehensive search of eight
databases took place to include articles published prior to November 2015. All
screening, data extraction and analysis procedures were performed in duplicate or
with multiple quality checks and consensus methods throughout. A narrative
synthesis approach informed the study analysis.

Results

Twenty papers met the inclusion criteria. Remediation outcomes of WBA
interventions have not been established within the existing literature. The
identification of general underperformance is not supported by the use of stand-
alone, single-assessor WBA events although areas of underperformance requiring

attention may be identified. Multisource feedback (MSF) tools may potentially facilitate identification of generalised underperformance.

Conclusion

The extent to which WBA can be used to detect and manage underperformance in postgraduate trainees has not yet been established although evidence to date suggests that multi-rater assessments (i.e. MSF) may be of more use than single-rater judgments (e.g. mini-CEX).

Introduction

The research problem

Determination of competence in postgraduate medical education is complex and the demand for better accountability in the assessment of performance standards and ensuring patient safety and quality of care continues to grow. One of the key challenges facing medical educators is the identification and remediation of underperformance. Almost two decades of research have attempted to determine whether the implementation of workplace-based assessment (WBA) can provide accurate, informative and learning-oriented judgments. While it appears that WBA does not appear to influence changes in practice among the general medical population[1-3] we do not know whether or how these assessments can assist in identifying poor performance (diagnostic assessment) or in remediating or changing practice among the subgroup of underperforming trainees or those at risk of underperforming. In this current era of outcome-based education and its reliance on valid and meaningful work-based and in-training assessment methods and practices, we aim to add to the evidence for the use of WBA in the specific context of underperformance.

Background

Workplace-based assessment (WBA) was introduced with the aim of providing trainees with observation-based feedback on their performance in a real-time work-based setting.[4, 5] In 1995 the Mini-Clinical Evaluation Exercise (Mini-CEX) was designed to specifically address emerging research at that time that trainees were rarely observed in clinical practice and even less frequently provided with meaningful feedback.[4] The implementation of WBA in postgraduate medical education and training programmes has consisted of various combinations of tools designed to address observation and feedback on for example, practical, technical communication and judgment skills including the mini-CEX, direct observation of procedural skills (DOPS) and case-based discussion (CbD).[6] Over time the implementation of these tools has varied widely and a debate now exists as to their

main purpose and role i.e. as an assessment of *performance*, or an assessment for *learning*.

This debate has been fuelled by a number of recent studies aiming to determine the best use of WBA. Firstly, evidence now suggests that WBA tools do not perform well as both summative and formative assessments[7, 8] in the main part due to confusion among users as to the main aim of the assessments.[9, 10] Secondly, the number of, for example, mini-CEX assessments required to make a reliable summative judgment is 8-10[11] which, in busy clinical settings is becoming less and less feasible and acceptable. Thirdly, emerging research on rater variability in assessment – including the complex and multi-faceted cognitive, social and psychological origins of this variability - has also raised questions as to whether WBA can reliably and validly be used to judge performance in a workplace setting.[12-16]

The identification of underperformance in postgraduate medical education is a complex challenge for educators. The literature continues to provide consistent evidence that the delivery of negative feedback is a significant concern even for experienced educators.[17] There is also continued evidence that many educators do not understand the formative - and therefore low-stakes - purpose of WBA[10, 18] potentially adding to fears of making erroneous or inaccurate judgments, balanced by the responsibility of ensuring that patients are safe in the care of their of trainees.

The development of the multisource feedback (MSF) tool often used in WBA, the Team Assessment of Behaviour (TAB) was explicitly designed to 'screen' for underperformance.[19] However, a single functional definition of 'underperformance' in this context is difficult to elicit. The TAB, like other forms of MSF, requires feedback from multiple team members about the trainee's performance within that team, and therefore focuses not only on clinical and technical proficiency, but on communication and ability to work as part of a team.

It has not yet been established whether or not a single assessment event detailing an area of performance that requires improvement is indicative of

underperformance as a whole. To date, the evidence suggests that the greatest impact of WBA lies in providing observation-based feedback but the impact or effectiveness of these tools on changing behaviour as a result of this feedback appears to be limited; however, the interpretation of the findings of a number of systematic reviews[1-3, 20, 21] is limited by a number of factors.

Firstly, studies that use changes in performance as the main outcome measures have interpreted this change in practice as 'evidence' of learning; while this functional definition is useful to an extent, it is limited and does not include the possibility of a learning effect from affirmation of good practice.[22] Secondly, the reviews we found did not look at whether or not a change in practice was required prior to the WBA episode or intervention; doctors performing at expected levels may not require as many changes to their practice as those deemed to be underperforming.

The impact – of lack of impact – of WBA on changing practice also needs to be considered in light of contemporary feedback literature, specifically studies addressing how trainees perceive and decide to act upon feedback. A recent series of studies addressing this issue provides some interesting insights into how trainees process feedback; using a regulatory focus theory to explore this question, Watling et al [23] attempted to understand the complexity and influence of a 'promotion' or 'prevention' focus on acceptance or denial of feedback. A key feature of other studies in this programme of research also highlighted the importance of feedback culture[24-26] and the credibility judgments that a learner makes about the feedback provider before deciding on the usefulness or relevance of that feedback.[27]

With the emergence and adoption of outcome-based education models, including competency-based medical education (CBME) in postgraduate medical education, comes a dependence on robust, longitudinal and continuous low-stakes assessment tools and methods that will aim to assist in tailoring learning and development to an individual trainee's needs and achievement of pre-defined programme outcomes. It is now more important than ever that the practice of WBA can therefore identify areas of strength, areas for improvement and allow for better

overall judgments of trainee achievement and performance throughout the continuum of their education. Early identification – and/or remediation - of underperformance is a key goal of these education models; the question for our review therefore remains as to whether WBA tools and the methods by which they are implemented can assist in identifying (potentially) underperforming trainees and whether or how WBA may also assist in remediation of this underperformance.

Review objectives

The aim of this review is to systematically evaluate the existing WBA literature to answer two overarching research questions:

1. Can workplace-based assessment be used to identify and remediate underperforming postgraduate medical trainees?

2. What features or implementation conditions of WBA tools specifically contribute to their usefulness in identifying or remediating underperformance among postgraduate medical trainees?

In light of a number of issues we identified prior to the review, the team agreed a series of definitions and terms for consistency throughout the study:

1. ***Underperformance*** within a clinical context is inconsistent within the literature and terms are often used interchangeably. The most contemporary (2013) definition provided within a UK-based study defines the underperforming trainee as 'requiring intervention beyond the normal level of supervisor-trainee interaction'.[28] While this definition does not classify the root cause of the trainee's difficulties it provides an overarching articulation of a trainee who is not currently meeting the expectations of their training level and we decided to use it as a reference to terms including 'the trainee in difficulty', 'the difficult/problem trainee' and 'the trainee in trouble'

2. ***Workplace-based assessment*** was defined as any assessment tool or method designed to provide feedback on performance and inform improvement in practice in a clinical setting and included (but was not limited to) tools such as:

    a. Mini-CEX

    b. DOPS

    c. CbD

    d. OSATS (Objective structured assessment of technical skills)

    e. Multi-source feedback (MSF) was used to refer to various tools designed to collect evaluations of performance by multiple assessors which is then collated and discussed with the trainee by a single facilitator. The tools in use include the Mini-PAT (mini-peer assessment tool) and TAB (team assessment of behaviour) and other formats referred to as 360° feedback

3. ***Postgraduate medical trainees*** are post-registration doctors enrolled in a training programme in medicine or surgery (e.g. resident, trainee, doctor-in-training, non-consultant hospital doctor)

4. ***Remediation*** is 'the act or process of correcting a deficiency' as described by Cleland et al (p e185).[29] This particular definition was chosen for our review as it links closely with the purpose of formative assessment, which is to provide information on performance strengths and deficiencies, to provide a structure for feedback and guidance on improving performance

5. The ***features*** or ***factors*** of workplace-based assessment tools we expected would be described included:

    a. The WBA rating systems and feedback structures

    b. WBA methods of use including such considerations as whether they are used routinely or in the case of suspected underperformance, if multiple tools are used and if one or many encounters were used in identifying or remediating performance-related issues

Methods

The study methods followed the BEME-approved study protocol,[30] the PRISMA guidelines for the reporting of studies[31] (Appendix 1) and the STARLITE standards for reporting literature searches[32] (Appendix 2).

Search Strategy

A detailed MEDLINE search strategy was developed in collaboration with a PhD-level information scientist during development of the study protocol (Appendix 3). Following acceptance and publication of the study protocol[30] the search strategy was reviewed by a second information scientist from a BEME International Collaborating Centre (BICC).

A third university-based information scientist collaborated with the lead author (AB) on the adaptation of the search strategy to the eight individual databases used for the study (MEDLINE, Science Direct, PsycInfo, Australian Education Index, British Education Index, ERIC, CINAHL and EMBASE) and also provided a 'tracking tool' as a search log for each database (Appendix 4). During the search process, a number of electronic database search alerts were also set up to identify any emerging research published prior to the analysis phase of the review.

The Cochrane database of systematic reviews, the BEME published reviews, and conference abstracts for AMEE (Association of Medical Education in Europe) and ASME (Association for the Study of Medical Education) from 2010-2015 were also searched along with specific medical education journals including *Medical Teacher*, *Medical Education*, *The Clinical Teacher* and *Advances in Health Sciences Education*.

Inclusion and exclusion criteria

The inclusion and exclusion criteria are presented in Table 1.

Table 1 Inclusion/exclusion criteria

	INCLUSION CRITERIA	EXCLUSION CRITERIA
Population	• Postgraduate medical trainees • Postgraduate surgical trainees	• Non-medical trainees • Medical students (undergraduate and graduate-entry programmes) • Studies not involving humans • Studies in medical areas **not** related to humans (e.g. veterinary studies) • Studies not involving physicians
Intervention	• Workplace-based assessment tools e.g. Mini-CEX, DOPS, case-based discussion, OSATS	
Outcomes	• Studies that described /reported outcomes related to identification/remediation	
Research Design	• No restriction for study design was applied	• Studies that did not report primary data • Reports published only in dissertation format

We included all studies that described or evaluated the use of workplace-based assessment in the following contexts:

• Routine or targeted use of WBA

• Trainee-led or trainer-led WBA

• Single or multiple WBA events

• Use of WBA as part of a wider programme of assessment or in the context of a range of assessment evidence

• Management or remediation of underperformance for knowledge, skills and attitudes

• Presence/absence of facilitation and/or written or verbal feedback

Evaluation of study outcomes

The primary outcomes of the review were those perceived to be resultant from the use of a workplace-based assessment intervention at the individual (trainee) level, practice level (*e.g.* change from non-routine to routine use of WBA by the training body or institution) or system-level (*e.g.* deanery-wide implementation of a new tool) (Table 2). Other outcomes included the conditions under which the use of WBA is most useful in identifying or remediating underperformance and, where possible, the features of WBA tools, or factors in using WBA, that are most likely to contribute to successful remediation of underperformance. Educational outcomes were organised using Kirkpatrick's framework of educational outcomes using Barr's adaptations for medical education research[33] and adaptations by Steinert et al[34] that subdivided Level 3 into self-reported (3a) and observable (3b) changes in behaviour (Appendix 5).

Table 2 Study Outcomes

Outcome	Example
Individual-level	• Number of trainees identified as poorly performing through the use (either routine or targeted) of a WBA process • Progression/remediation statistics • Changes in trainee performance (knowledge, skills, attitudes *etc.*) • Trainee satisfaction
Practice- level	• Changes in implementation methods, *e.g.* non-routine to routine • Implementation of new/differing WBA tools
System-level	• Changes in system-wide implementation of WBA tools or methods *e.g.* throughout a deanery

Assessment of methodological quality

a. Observational studies

We chose to use the methodological quality assessment described by Buckley et al[35] for the evaluation of observational studies. Each criterion was independently rated as 'met', 'unmet' or 'unclear'. This framework suggests that in order to be deemed of high quality, studies should meet a minimum of seven of these quality indicators. Recent guidelines suggest that reporting of ethical issues should consider both ethical approval for a study and issues of informed consent separately.[36] Where this is not reported the guidelines suggest that this be deemed 'unclear'. We therefore modified the BEME criterion ('are all ethical issues articulated and managed appropriately?') to these specifications.

b. Qualitative studies

We had intended to evaluate any identified qualitative studies using the COREQ (consolidated criteria for reporting qualitative research) guidelines.[30, 37] However, this checklist was designed to facilitate reporting of interview and focus-group based studies.[38] Of the two included qualitative studies in this review, one used mixed methods and it was agreed that the Critical Appraisal Skills Programme (CASP) guidelines[39] for the reporting of all qualitative studies would be more appropriate in this instance.

Data extraction and coding

A modified BEME coding sheet (Appendix 6) was developed by the review team and approved by the team's appointed BEME International Collaborating Centre (BICC). Data that were extracted in the Excel spreadsheet from the studies included:

- Study design, presence/absence of a conceptual framework
- Population and setting including training programme, year in training
- WBA 'intervention' tool characteristics (including rating scales) and method of implementation

- Conditions of use i.e. routine/targeted and any specified implementation factors
- Evaluation and outcomes of the study including educational impact, identification and/or remediation outcomes

All included articles were independently coded by two of the investigators (AB and RG). A coding pilot was performed in which both coders independently reviewed a set of two papers and a third reviewer, unrelated to the project and working in another BICC performed a quality check of the data and reported no major discrepancies in data extraction between the two coders.

Results

Selection of papers

The database searches identified 7067 papers. Following de-duplication this resulted in 6261 papers for screening. All searches were imported into EndNote X7. A flow chart of the selection process is detailed in Fig 1.

The initial screening was completed by AB; at the start of the process, a quality check was performed in which AB and RG independently screened the titles and abstracts of the first 836 citations and performed an inter-rater reliability calculation. Current practice suggests that Cohen's weighted kappa and percentage agreement both have strengths and limitations and more than one determination of agreement should be used;[40] we therefore performed a percentage agreement (99.04%) along with a weighted kappa (Appendix 7) which was 0.641 indicating moderate-to-good agreement but also cannot rule out the role of chance in this agreement statistic. All disagreements at this stage (8 of 836 citations) were resolved by discussion.

Of the 6261 papers screened, 6059 were excluded on the basis of title and abstract. The full text of 202 papers was retrieved and AB performed an initial screening, compiling those papers into three files: *exclude*, *include* and *for team discussion*. At this stage 169 papers were excluded, 16 were included and 17 were identified as requiring further discussion by the review team. RG performed a second quality check of all three sets and there was complete agreement.

Due to the international make-up of our review team, we devised a novel 'voting' system for the papers requiring further consideration. Each paper was reviewed by a minimum of 3 team members and a 'voting spreadsheet' was compiled to identify (non)agreement on inclusion (Appendix 8). During this process there was complete agreement on the exclusions of nine papers. For the remaining eight papers, team discussion took place via email. Twenty studies were included in the final review.

All included papers were journal articles and ranged in time from 2000-2015. The study designs included evaluative and retrospective studies, with the majority taking place in hospital settings.

Fig 1: Flow diagram of search and selection process
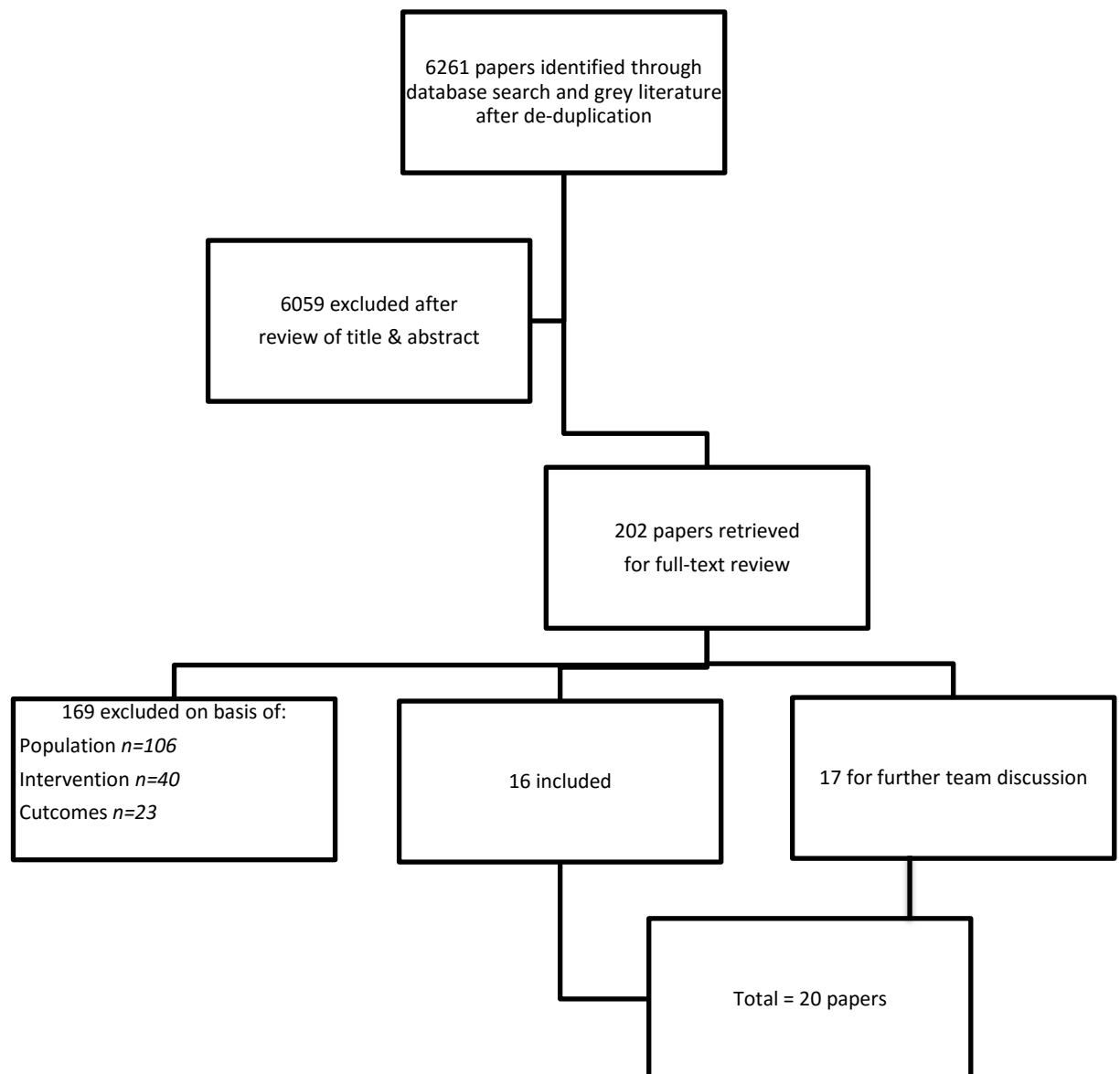
Review Findings

Overview

Of the 18 quantitative studies included, multi-source feedback was the WBA 'intervention' in 13 studies (Table 6); six of these studies specifically used the mini-TAB format (Table 7). Two cross-sectional studies surveyed programme directors about the methods used to identify trainees in difficulty[41, 42] and one study attempted to determine whether scores in a number of WBAs could predict underperformance. One study considered OSATS only[43] and the minicard direct observation tool was evaluated by Donato et al.[44] A new pharmacotherapy-structured clinical observation (P-SCO) tool was implemented and evaluated by Young et al.[45] The qualitative studies included an evaluation of anaesthetic trainer and trainee perceptions of the mini-CEX[46] and participant experiences of a new model of facilitation of multisource feedback[47] (Table 8).

Methodological quality

The methodological quality of all included studies was independently evaluated by two reviewers (AB and RG) and minor differences were resolved by discussion. Buckley et al[35] suggested that of the 11 BEME quality indicators, studies meeting seven or more of the criteria may be considered of 'good' methodological quality.

Of the eighteen observational studies included in this review, ten studies met this requirement and another study met nine of the suggested criteria. However, it is worth noting that two specific methodological evaluation judgments should be considered in the context of the time at which the majority of studies took place and which may impact on the overall impression of methodological quality:

1. None of the studies included in the review provided a pre-identified conceptual or theoretical framework. The inclusion of such a critical 'lens' through which a study's results can be interpreted or analysed is a relatively recent development in medical education research.

2. We also attempted to identify ethical issues associated with both informed consent and ethical approval. Until relatively recently, medical education research was generally considered exempt from institutional approval as it was not perceived to involve 'risk' to the participants. Complete reporting of ethical issues was a limitation of all 18 quantitative studies. While a number of studies reported prospectively obtaining research ethics committee approval, informed consent was not evident in 10 studies where trainee data or participant information were analysed.

a. Methodological quality of quantitative studies

Two of the quantitative studies included in our review were prospective in design[19,48] and three studies were cross sectional (Appendix 9).[41, 42, 49] Of the 13 remaining studies, seven were retrospective while the timeline of the intervention (vs the research study) was unclear in the other six studies. A number of studies reported conclusions that were not fully supported by the data emerging from the studies; in these studies the data appeared to be retrospectively reported and would have been enhanced by pre-intervention consideration of study outcomes. While study designs and data collection methods were generally appropriate to the stated research question, the strength of the findings was limited by the absence of before-and-after interventional studies. Two of the included cross-sectional studies[41, 42] were reliant on participant recall of the number of trainees identified as being 'in difficulty'. Unfortunately these data were not triangulated with any documentary evidence which contributed to the limitations of the study conclusions. Lack of triangulation in general was a limitation of 12 of the 18 quantitative studies.

Although none of the included studies were randomised trials, we included a specific quality indicator addressing whether the study authors had attempted a risk-of-bias assessment *and/or* included a statement of researcher positionality (modification of the BEME quality indicator No.5[35]). Only a single evaluation study[45] explicitly addressed risk of bias and stated that the raters involved in the study had

not been blinded to the study hypothesis. A large-scale study by Archer et al[50] would also have been potentially enhanced by an assessment of risk-of-bias inherent in the study design in which MSF raters were not blinded to the fact that the doctors under assessment had already been referred to the National Clinical Assessment Service with suspected issues of underperformance.

Table 3: Included quantitative studies (not including studies of Team Assessment of Behaviour)

Study	WBA 'intervention'	Journal of publication	Country	Design	Retrospective /prospective	Population	Outcomes of relevance to BEME review	Methodological quality assessment
Archer et al, 2011[50]	MSF - SPRAT	Medical Education	UK	Cohort study	Unclear[a]	All doctors referred to NCAS including trainees	1. MSF and patient rating scores compared to reference 'normative' group. • NCAS-assessed doctors scored lower on MSF than 'normative' group but not on patient ratings	7
Black and Welch, 2009[51]	MSF - mini-PAT	The Clinical Teacher	UK	Cohort study	Retrospective	FY Doctors in difficulty	1. Number of doctors identified by deanery as 'in difficulty' using mini-PAT alone (16.6% = 10/60)	5
Brown et al, 2008[42]		Laryngoscope	US	Cross-sectional study		Programme Directors	1. links between programme design reporting of underperformance or remediation 2. programmes with remediation mechanisms in place, or with formative feedback practices more likely to report incidence of 'resident remediation' in last five years	6
Chipp et al, 2011[52]	MSF as prep course	Medical Teacher	UK	Descriptive paper	Unclear[a]	FRCS(Plast) candidates, n=9	1. MSF consultant scores compared with actual exam scores – 'accurately' predicted scores in 6 of the 9 trainees	7
Donato et al, 2015[44]	Minicard direct observation tool	Journal of Graduate Medical Education	US	Retrospective cohort analysis	Unclear[a]	Internal medicine residents year 1-3	1. 9% of 7345 individual ratings deemed unsatisfactory 2. 1st year residents most often rated as 'good' vs 3rd yr residents most often rated as 'excellent' – no	7

								indication of statistical significance of difference between groups	
Hesketh et al, 2005[53]	360° 'diagnostic screening tool' as component of PHAST assessment system. 42-item questionnaire	Medical Teacher	Scotland	Descriptive / evaluation study – multiple methods	Retrospective	PRHOs (pre-registration house officers)	1. Identification of underperformance (Requires help/attention = Rh\A) • 65 (27%) trainees had 1-10 RH/A • 13 (6%) trainees had 11-20 RH/A • 3 (<2%) trainees - >20 RH/A 2. Actions arising varied: • <4 RH/A: trainees received satisfactory overall rating • 4-10 RH/A: mixed actions, some repeated PHAST +/- individualised support	7
Hiemstra et al, 2011[43]	OSATS	Canadian Journal of Surgery	Netherlands	Single group, no comparison	Unclear[a]	Ob/Gyn trainees, PGY 4, n=9	1. Learning curves of OSATS • identification of underperformance based on 'threshold score' of av of 75% on each of the six domains • 2 trainees never achieved threshold score	7
Mitchell et al, 2011[54]	CbD, mini-CEX, DOPS, mini-PAT	Medical Education	UK	Observational study	Retrospective	FY1 & 2	1. Compared WBA scores of trainees already 'flagged' as in difficulty with others (WBAs used numerical and narrative ratings) • Mean mini-CEX and CbD scores were lower for trainees in difficulty but 'weak predictive value' • No statistically sig differences in mean mini-PAT and DOPS scores	4

Vivekananda-Schmidt et al, 2013[55]	MSF	Medical Education	UK	Cohort study	Retrospective	11483 forms from 933 drs; 4777 with free-text comments (42%)	1. 513 forms had at least one 'below av' rating; 56% (286) had free-text comment 2. comments generally multi-domained and assessor-centred (vs trainee-centred) and did not provide evidence/examples of underperformance or how to change practice	6
Warm et al, 2010[56]	MSF	Journal of Graduate Medical Education	US	Descriptive/ evaluation study	Unclear[a]	Internal medicine residents	Outcomes were trainee rankings on each domain of each component of the MSF process. Identified poor performance relative to peers, and compared to quality data	7
Yao and Wright, 2000[41]	N/A	Journal of the American Medical Association	US & Puerto Rico	Cross-sectional study		Internal Medicine Programme Directors	• 6% PDs felt they had no 'problem residents' • 8% stated mini-CEX identified 'problem residents' 'half of the time or more frequently' • No WBA appeared in list of 'interventions' used to deal with 'problem residents' • 16% Community PDs responded that mini-CEX identified 'problem residents' half of the time or more vs 4% of hospital PDs	6
Young et al, 2011[45]	P-SCO Pharmacotherapy-Structured Clinical Observation tool	Academic Psychiatry	US	Evaluation study	Prospective	3rd year Psychiatry residents	1. P-SCO scores vs global assessment ratings: • Trainees received an average of 1.7 P-SCO ratings 'below expectation' compared with 0 on global assessment	9

[a] Timeline of study (vs intervention) unclear

Table 4: Included Team Assessment of Behaviour (TAB) studies

Study	Journal of publication	Country	Design	Retrospective /prospective	Population	TAB format	Outcomes of relevance to BEME review	Methodological quality assessment 'score'
Bullock et al, 2009[57]	Medical Education	UK	Cohort study	Retrospective	Senior House Officers (2003); FP 1 and FP2 trainees (2005)	4 domains, 3-point rating scale (no concern – you have some concern – you have a major concern)	1. Number of 'concerns' documented by assessor groups = 6% in total 2. rater differences between groups = SHOs and FY doctors more lenient than consultant and nurse assessors	7
Burford et al, 2010[49]	Medical Education	UK	Cross-sectional, survey study		Trainees, raters and supervisors	4 domains, 3-point scale (no details provided). Mini-PAT format = 16-items with 6-point rating scale	Compared perceptions of users of mini-PAT vs users of TAB on 5-point Likert scale statement 'the feedback provided by this form could identify a doctor in difficulty'. Overall mean = 2.99 Mean ratings: Trainees = 2.56 (mini-PAT) vs 2.89 (TAB) Supervisors = 2.9 (mini-PAT) vs 1.05 (TAB) Raters = 3.06 (mini-PAT) vs 0.98 (TAB) Statistically significant differences not provided	7
Ellul et al, 2012[58]	The Clinical Teacher	Malta	Cohort study	Retrospective	Foundation programme trainees Year I and II	4 domains, 3-point scale (no concern – some concern – major concern)	40 assessments on 18 doctors (FY1 x 12, FY2 x 6) had any concern. Of the 18 drs, 2 (10.8%) had a concern on	6

							more than one MSF. No sig diffs between nurses and doctors (assessors) in identification of concerns	
Hassell et al, 2012[19]	Postgraduate Medical Journal	UK	Cohort study	Prospective	FY2 trainees in 13 hospitals	4 domains. 4 sets of scales i.e. 3-, 4-, 6- and 9-point scales	• Outcomes related solely to differences in ratings between 4 formats of TAB • Longer scales associated with fewer ratings of concern; TAB6 and TAB9 trends towards more trainees rated 'above expectation' than TAB4. • No significant differences in survey of assessor perceptions that forms were 'very effective' in identifying underperforming trainees (51%, 48%, 48%, 43%)	7
Whitehouse et al, 2007[48]	Medical Teacher	UK	Description/ evaluation study	Prospective	Senior House Officers	4 domains, 3-point scale (no concern – some concern – major concern)	• 109 SHOs (63.74%) with 'no concern' in all domains • 62 trainees with some concern by either 1 or more assessors • 21 SHOs with concerns x >1 assessor • Six ed supervisors felt that concerns they had were not identified during the process	7

| Wood et al, 2006[59] | Medical Teacher | UK | Description/ evaluation study | Unclear[a] | All training years O&G | 4 domains, 4-point scale (needs serious attention – progress needed – fine, no problem, outstanding). Trainees did not choose assessors | • 405 forms contained ratings in lower two categories, deemed a 'fail' but no mention of number of trainees to which this equates | 6 |

[a] Timeline of study (vs intervention) unclear

Table 5: Included qualitative studies

Study	WBA 'intervention'	Journal of publication	Country	Design	Population	Outcomes of relevance to BEME review	Methodological quality assessment (CASP) 'score)
Sargeant et al, 2011[47]	MSF	Medical Teacher	UK	Qualitative study	GP trainers and trainees (x 13 & 13)	1. Educational outcomes included trainers' reporting of observable changes in behaviour but these weren't supported by any specific evidence of change. 2. Trainees' education outcomes were at level 3a - self-reported changes in behaviour	8/10
Weller et al, 2009[46]	Mini-CEX	British Journal of Anaesthesia	Australia and NZ	Qualitative/ Mixed methods evaluation study	Anaesthetic trainees and specialists	Trainers noted that feedback was easier to give with mini-CEX but remained unwilling to write down any areas of concern	7/10

b. *Methodological quality of qualitative studies*

The methodological approaches to both qualitative studies and their study designs were generally appropriate to the articulated research questions (Appendix 10).

The methodological quality of the study by Sargent et al[47] was good, with the study meeting eight of the ten CASP criteria. The main issue of concern to the validity of this study was the lack of consideration of the influence of the researchers on the data collection and analysis process, given that they had also acted as facilitators of the workshops around which the study was based. The interview topic guide was not provided and therefore it is unclear as to whether the study methods were entirely matched to the research question.

The study by Weller et al[46] also met seven of the ten CASP criteria; the main issues limiting the quality of this study involved lack of clarity around recruitment and what, if any, influence the researchers had on the data collection and/or analysis process. While the interviewer-participant relationship was articulated, the relationship between the study gatekeeper and the invitees was unclear.

Review outcomes

1. Trainee-level outcomes

Using the modified Kirkpatrick's hierarchy of educational outcomes, five of the studies reported outcomes at *Level 1a* i.e. learner reaction to the educational intervention.[46, 48, 49, 52, 53] Outcomes at *Level 2a* (i.e. evidence of change in skills) were reported by Hiemstra et al[43] by developing learning curves to plot the trajectory of OSATS ratings and achievement of pre-defined levels of 'good' practice. Self-reported changes in trainee's clinical practice (*Level 3a*) were reported by Sargeant et al[47] in which GP trainees were interviewed about the impact of a new model of feedback delivery in the context of multisource feedback; however, as is the limitation of self-reported behaviours, this outcome was not verified by triangulation with any other performance data.

We also explored the data for evidence of any other impact of the WBA intervention on the individual trainee. In general it appeared that where outcome data were reported, trainees generally improved performance or progressed to the next stage of training. Black and Welch[51] reported outcomes for trainees identified as 'in difficulty' including those identified as underperforming using the mini-PAT. However these outcomes were reported for all trainees in difficulty, regardless of the mechanism of identification. Therefore of the 10 trainees identified as underperforming using the mini-PAT, the specific issues identified and specific outcomes for those ten trainees were not explicitly reported.

Chipp et al[52] reported correlations between the MSF intervention (in this case feedback from supervisors and peers on technical performance) and FRCS (Plast) scores but specific improvements in knowledge or skills were not reported. It was also unclear whether participation in the WBA intervention *improved* FRCS performance. This was a very small study (nine participants) therefore the results in terms of outcomes will require further large-scale investigation before those outcomes can inform practice- or system-level changes.

Brown et al[42] reported on a survey of programme directors in which the presence or absence of a remediation mechanism was explored. In their conclusion they reported that programmes with an established remediation programme were more likely to report the identification of trainees requiring remediation, however the outcomes (and descriptions) of these remediation interventions was not provided.

In another survey of programme directors, Yao et al[41] reported that participants 'estimated' the residency completion rate among problem residents as 57%, with 18% requiring additional time but completing the programme, 9% moving to another similar residency programme, 10% moving to a difference residency programme and 4% leaving medicine. These study findings are however, limited by reliance on recall and were not triangulated by any documentary evidence to support the findings. It was also unclear as to how many of these 'problem residents' had been identified using WBA or if WBA was a feature of any of the remediation interventions.

The development and implementation of an assessment system incorporating multisource feedback was described by Hesketh et al.[53] They reported that of the trainees given a rating of 'requires help/attention' (Rh/A) on any part of the assessment system (including a 360° feedback), the resultant outcomes trended towards an overall satisfactory evaluation where there were less than 4 Rh/As. In the case of trainees with 4-10 Rh/As, educational supervisors differed in their treatment of the case with some trainees deemed to require a repeat evaluation and others progressing, while trainees with up to ten Rh/As were all required to repeat their evaluation. The overall outcomes for this group, along with any remediation interventions, were not reported.

2. Practice-level and system-level outcomes

None of the studies reported changes in the practice of using or implementing WBA or in system-level outcomes. Seven studies described the development and/or implementation of a WBA tool or methods[44, 45, 47, 52, 53, 56, 59] however, it was unclear as to whether the findings or outcomes resulted in the adoption or dismissal of these new tools or methods.

Synthesis of findings

In attempting to answer our original research questions we have synthesised the review findings to discuss firstly, the use of WBA in identification of underperformance and secondly, its use in remediation of underperformance among postgraduate medical trainees. Although the variation in study purposes, designs, outcomes and implementation among the studies precluded a meta-analysis, we have instead attempted to provide a narrative synthesis of themes emerging from the included papers in the context of these original research questions.

1. How can WBA be used to identify underperformance?

It appears from the studies included in this review that WBA may have a role in the identification of underperformance. While the optimal mechanisms and WBA implementation conditions to achieve this goal are not yet clear, the emerging themes presented here tend to support the longitudinal integration of regular, continuous low-stakes WBA as important determinant of overall performance among this cohort.

1. *Routine or targeted use of WBA*

In general the included studies reported on the routine use of a WBA tool or tools for all trainees within a specific cohort. Of the twelve studies that described the routine (non-targeted) use of WBA, nine referred to the routine use of MSF. The majority of the MSF events occurred once per year or once per six-month training post.

In a single study reporting the targeted or purposeful use of WBA, Archer et al[50] described the use of MSF for doctors referred to the UK National Clinical Assessment Service (NCAS) with suspected issues of underperformance. The study aimed to explore the ability of the tool to discriminate between doctors previously identified as underperforming and a 'normative reference group' of pilot

participants. While this study did reveal that the MSF scores for the underperforming group were significantly lower than those of the control group, two additional study design issues need to be considered in the interpretation of these findings.

Firstly, the assessors who completed the MSF were aware that the doctor being assessed had been referred to the NCAS. From the vast literature on assessor rating variability[13-15, 60] and evidence that assessors are reluctant to provide negative feedback in face-to-face situations[17, 18] this may have swayed assessors' ratings and potentially facilitated the delivery of negative feedback if they were reassured that the underperformance had already been identified by others.

Secondly, the NCAS provides assessments for all doctors, including trainees, but this study did not distinguish between trainees and non-trainees. It is therefore not possible to extrapolate their findings purely to the group of underperforming trainees. It would be interesting to further explore this study's data to ascertain differences in the discriminatory ability of the MSF tool between these two cohorts.

2. Trainee- or trainer-led WBA

Multisource feedback (MSF) is characterised by the collection of feedback from multiple assessors. In general, these assessors are identified by the trainee and supervisor together. In our review, MSF was implemented in this way by all authors with the exception of Wood et al.[59] These authors described the identification of assessors by supervisors only which they determined was justified in the context of the pilot phase in which this was implemented. In general, implementation of the routine MSF was either voluntary or mandatory but coordinated by the local educational supervisor or deanery lead.

None of the studies reported or described trainee-led implementation processes and methods therefore a comparison of the impact of either method on identification or remediation use of WBA could not be established.

3. Single or multiple use of WBA tools

Studies included in the review generally reported on single WBA events and therefore on immediately related outcomes. This is in direct contrast with the premise of WBA which supports the use of multiple low-stakes assessments and with current evidence that suggests that a single WBA event is not, by itself, a reliable judgment of overall trainee performance. However, the use of a single WBA event to identify and provide feedback on *areas* of underperformance has not yet been established.

Of the 13 multisource feedback studies identified in this review, there were no comparison studies of the use of single-vs-multiple events and their impact on outcomes. It was also not possible to determine from the study by Mitchell et al[54] whether programmes using Mini-CEX, DOPS and case-based discussion had implemented these WBA tools according to research guidelines for the number required for good reliability e.g. in the case of mini-CEX, whether a minimum of 8-10 assessments were recorded.[11] This retrospective study looked at 'mean' scores for each assessment type (mini-CEX, DOPS, CbD, mini-PAT) for each trainee by converting the narrative ratings (expectations-based) into scores of 1-6 and retrospectively compared these means for two groups i.e. trainees already 'tagged' as in difficulty and those who had not been tagged. While 'associations' were noted between lower mean scores on mini-CEX and CbD in trainees already tagged as in difficulty, there was little evidence of predictive ability of WBA to identify trainees in difficulty using ROC (receiver operator characteristic) curves.

The study did not provide information as to how underperformance had been identified among this group and did not evaluate how many assessments were performed per trainee to generate this mean. The use of a single 'mean' score also limited the ability of the tool to identify underperformance as this may not pick up on subtle 'dips' in performance as opposed to trends seen over time.

4. *Use of WBA as part of a wider programme of assessment or in the context of a range of assessment evidence*

The use of WBA as part of an overall system or programme of assessment was not a key feature of the studies included in this review. However, a single study by Hesketh et al[53] described the use of MSF as part of a PHAST (pre-registration house officer appraisal and assessment system) system of assessment for pre-registration house officers (this study took place prior to the establishment of the 2-year Foundation Programme in the UK). The 360° assessment tool was completed by four raters twice in the year using a 4-point narrative scale ('excellent – good- satisfactory - requires help/attention'). The paper provided details of the implementation of this 360° assessment alone and did not compare the results to the entire portfolio of evidence generated within the PHAST system and therefore could not be compared to using the MSF tool alone.

In another study, internal medicine residents participating in a year-long ambulatory clinical attachment were assessed at two points in the year using MSF along with clinical quality data, patient ratings and knowledge-based test scores and ranked relative to peers for each component.[56] The authors were able to identify poor performance relative to peers and compared to quality data, but it is not clear from the study whether this ranking system was more or less effective than using MSF alone to identify underperformance, although there were a number of highlighted areas of inconsistency between MSF scores and other values.

5. *Presence/absence of facilitation and/or written or verbal feedback*

The addition of verbal or written feedback on the identification of underperformance appears to impact positively on the quality of the WBA. Nevertheless, the impact of that feedback on detection of underperformance appears to depend to a large extent on the quality and/or specificity of that feedback.

A qualitative study by Sargeant et al[47] explored the impact of an 'ECO' (emotion, content, outcomes) model of facilitated feedback on trainer and trainee

perceptions of the MSF process. As described above, the impact of the intervention was only evaluated at the level of self-reported changes in practice among trainees; nevertheless all participants (GP trainers and trainees) were positive in their evaluation of the model. The study did not, however, explore whether this model was more or less useful than usual MSF processes in identifying or remediating underperformance.

The 'learner-centeredness' of written MSF feedback was analysed by Vivekananda-Schmidt et al.[55] Of a total sample of 11483 MSF forms, only 4777 (42%) contained any free-text comments. Using a content analysis approach, the authors determined that where feedback was provided, this generally tended to be 'rater-centred', with an emphasis on the trainee's impact on the assessor's working life rather than on goal-oriented feedback for the trainee's development.
The authors also specifically analysed the 513 forms containing a 'below average' rating; of these, 56% contained free-text feedback despite explicit instructions that all such ratings should be accompanied by feedback. Given the lack of trainee-centred feedback in general, and specifically in trainees at risk of underperformance, the authors concluded that MSF may be of limited use in identifying or remediating underperformance where the feedback is not informative.

Young et al[45] explored the implementation of a new tool, the Pharmacotherapy Structured Clinical Observation Tool (P-SCO) for third-year psychiatry trainees and compared the written comments on the tool to those provided on the comparator tool, a global rating scale. Their results showed that assessors were more likely to provide specific comments on the P-SCO compared to the global rating scale, providing 2.6 times more affirmatory feedback and 5.3 times more corrective feedback. The P-SCO also identified more ratings of 'below expectation' than the global rating scale. The form specifically requested 'key feedback points, including what was done well and at least one task to work on' and all assessors had participated in a faculty development workshop prior to the implementation.

6. *Rating scales*

While MSF rating scales varied slightly throughout all 13 studies, one paper specifically addressed whether scale length impacted on the number of 'below expectation' ratings identified by assessors.[19] Using 4 versions of the Team Assessment of Behaviour (TAB) MSF tool, the authors reported trends towards fewer underperformance ratings using the longer versions of the scale. However the study design meant that the four versions of the form were used in four different training locations; there was no direct comparison of rating scales among a single group therefor the findings are limited in their generalisability.

7. *Rater variation*

It is widely accepted that rater variability is an important factor in any assessment of performance and this has been widely studied in the context of workplace-based assessment.[12-15] Our review included one such study which looked at whether different rater groups were more or less like to identify a trainee in difficulty.[57] While they determined that some assessors were more likely to be more lenient that others and the study reported concern ratings, the impact of this rater variability specifically on detection or remediation of underperformance was not fully explored.

2. How can WBA be used to remediate or manage underperformance?

This aspect of the review remains unanswered. We did not find any studies that specifically looked at the impact of WBA on remediating or managing underperformance. In the single included study that attempted to explore associations between remediation processes and trainee outcomes, Brown et al[42] surveyed 100 otolaryngology programme directors to determine which assessment tools were in place across all programme. This study found some weak correlations between the provision of formative feedback and identification of underperformance. Where programmes had a remediation mechanism in place, they were more likely to identify underperformance but the study could not identify whether having the mechanisms in place first allowed for better identification *or* if issues related to identification of underperformance necessitated the development of remediation mechanisms.

The usefulness of these findings were limited by the study design; this was a survey of programme directors in which data collection relied upon on recall and was therefore subject to recall bias as there was no triangulation with documentary evidence of trainees in difficulty. The self-reported feedback mechanisms may or may not reflect actual practice.

Discussion

Over the past twenty years since the inception of the mini-CEX[4] a vast body of literature has emerged on workplace-based assessment. While a number of previous systematic reviews have failed to unearth definitive effectiveness of WBA in changing practice[1-3] we recognised that in a group of well-performing doctors, change may not always be necessary and we therefore attempted to address this question within the specific context of underperformance.

The aim of this BEME review was to systematically examine whether or not WBA is of use in identifying or remediating underperformance and if so, to identify which tools or features of those tools (or their implementation) may contribute assist in this endeavour. Our results have allowed us to explore this previously under-researched topic from a narrative perspective and to understand the limitations of current research in contributing to this important conversation in postgraduate medical education. While it appears that the routine integration of some WBA methods and tools may assist in identifying *areas* of underperformance, its use in identifying trainees who are generally underperforming is not yet clear. Although this is due to multiple factors, including the implementation challenges and variations encountered by training bodies and institutions, the absence of well-designed interventional studies also limits our ability to answer this question definitively.

In this section we present an overview of the study findings and their implications for current and future practice, along with considerations for future research to better understand the role of WBA for underperforming trainees.

Main review findings

WBA and underperformance

Our search strategy uncovered a number of studies reporting underperformance data; in many cases the authors proposed that this in itself provided evidence of the ability of the tool to identify underperformance but these data were generally not supported by any other sources. In two studies that compared WBA ratings of performing and underperforming trainees[50, 54] the strength of association provided in the results was limited by the study design. In both cases, the group of underperforming trainees had been identified or flagged as underperforming by other means which were unclear, possibly by 'expert opinion'. In particular, the study by Archer et al,[50] while providing statistically significant differences in ratings between the two groups, was limited by a design bias in which the MSF assessors for the underperforming doctors were not blinded to the fact that the doctor under assessment had already been identified as underperforming and had been referred to the NCAS for assessment.

On reviewing the included studies as a whole, it is not possible to definitively articulate whether WBA is effective in identifying underperformance among trainee doctors. The majority of studies reported the number of trainees or assessments that had a concern rating but the implications and outcomes of this detection were in general not provided and in very few cases triangulated by other performance markers. Nevertheless, we can make some observations on trends we found throughout the studies.

Multisource feedback was the most commonly used WBA tool used to report ratings of concern that appeared in this review and it is worth noting a number of elements of that tool – and its implementation – that may influence its ability to identify underperformance more than other tools.

Firstly, the MSF process requires input from a number of assessors. In this review, the majority of MSF interventions involved more than six assessors and trainees in

all except one study[59] were involved in choosing their assessors. The responsibility of providing ratings is cushioned by the fact that the feedback is collated from all assessors and is delivered to the trainee in this format. Therefore in contrast to, for example other single-rater WBA tools, an assessor may be more willing to provide below average or concern ratings. This may also militate against the bias effects of a single-assessor WBA in which rater variability may provide a threat to reliability and validity.[12, 15, 16, 60] In our review Bullock et al[61] noted differences in MSF ratings among different professional groups which are in line with other studies of rater variability. The collation of feedback from up to ten assessors may therefore provide a more reliable overall assessment of performance than a single-rater judgment.

Secondly, the majority of studies reported the routine use of the MSF once a year, or twice at most. Again in contrast to other WBA tools that document a single WBA event, the MSF assessment and ratings provide for general impressions of the trainee over time and do not focus on a single interaction. The supervisor or trainer therefore delivers collated feedback which may assist in delivering negative feedback if issues of concern have been identified by more than one rater.

While a subgroup meta-analysis of the team assessment of behaviour (TAB) tool could not been performed, it appears that the shorter versions of the form (using 3- or 4-point scales) appeared to have slightly better detection rates of underperformance than longer versions.[19] The limitation of this finding however, lies in the fact that the variations of the forms were used in different cohorts of trainees and therefore we cannot imply that for the same group, one type of rating is better than another.

It appears that MSF is generally implemented in the same manner across all training programmes, however the various MSF tools (including generic MSF, mini-PAT and TAB) have not been compared to each other and it would be worth considering whether one tool may be superior to others in its ability to detect or remediate underperformance.

WBA and remediation

It appears from this review that the use of WBA in remediation of underperforming trainees has not been thoroughly explored. Given the contemporary balance of evidence for the value of WBA as a formative (rather than summative) tool, its value in providing observation-based, learner-centred feedback should be considered. This is particularly important in the context of programmatic approaches to assessment in which it is evidence of multiple types of assessments, under multiple conditions, and using multiple assessors that will create the overall picture of the performing – or underperforming trainee.

Implications for practice

Although the findings of this review have few definitive or immediate implications for practice, we have attempted to bring together emergent themes and trends that support the use, for now, of WBA in identifying underperformance. Further research is required to determine whether certain tools (and/or their implementation methods) are better than others in detecting underperformance and this will require robust comparison-based study designs along with consideration of the interpretation of single-episode concern ratings as opposed to ongoing underperformance issues. It currently appears that of the WBA tools identified in this review, MSF may have the ability to detect underperformance more than other single-rater tools where the number and range of assessors is adequate to do so.

Recommendations for future research

As discussed above, there is an urgent need for comparison-based studies to determine which tools may be effective in identifying underperformance and the conditions under which this can be achieved. This research is of relatively urgent importance with the exponential growth of outcome-based approaches to medical education, approaches which are dependent on continuous observation and the provision of high quality, trainee-centred feedback.

One of the larger gaps in the literature we identified included a consistent definition and description of the *underperforming trainee* versus indicators of concern or specific *areas of underperformance*. It is important to determine how many concern ratings, or what patterns of WBA ratings may indicate trainees who are in difficulty. Validity studies, informed by contemporary understanding of the concept of validity[62] and the use of newer validity frameworks[7, 63, 64] should also be considered in determining the value of WBA in identifying and/or remediating underperformance.

Strengths and limitations of the review

The main strengths of this review included the breadth of expertise and experience of our research team. We were fortunate to have a number of experienced BEME and Cochrane reviewers who brought significant methodological strength and rigour to the development and execution of the review process.

The consultation with three information scientists at various stages of the review also allowed us to ensure that our search strategy and searches were systematic, complete, thorough and rigorously documented. We performed quality checks at all appropriate stages of the review and given the international make-up of our team, the inclusion of our email 'voting' and discussion process allowed us to ensure that the final set of review articles met our inclusion criteria.

Conclusion

Evidence for the use of WBA in identifying and/or remediating underperformance among postgraduate medical trainees has not yet been established. While this is partly due to the quality and focus of studies already published, it appears that the question of how useful WBA is for this group of postgraduate medical trainees has not been addressed in general. Nevertheless we hope that this review will be of use in designing focused programmes of research aiming to definitively determine the role and value of WBA for this specific postgraduate medical education group.

Authors' contributions

AB (review lead) conceived the original research question, performed the database searches, screening, data extraction, coding, and data synthesis/analysis, drafted the final paper and coordinated all revisions. RG advised and collaborated on all aspects of the study design and execution and performed second checks of all data screening, independently coded included papers and assisted AB in the data synthesis/analysis and write-up. YS advised on all aspects of the review design and execution, contributed to the data synthesis/analysis and the final paper write-up. A O'S contributed to the study design and final write-up. AS advised on, and contributed to, all aspects of the review design and execution including data synthesis/analysis and final paper write-up. MH advised on and contributed to all aspects of the review design and execution, including the data synthesis/analysis and final write-up. TH advised on all aspects of the review design and execution with particular focus on methodological approaches to the data synthesis/analysis and contributed to the drafting of the final paper.

Acknowledgements

References

[1]   Overeem K, Wollersheim H, Driessen E, et al. Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study. *Medical Education*. 2009;**43**:874-882.

[2]   Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;**341**.

[3]   Saedon H, Salleh S, Balakrishnan A, Imray C, Saedon M. The role of feedback in improving the effectiveness of workplace based assessments: a systematic review. *BMC Medical Education*. 2012;**12**:25.

[4]   Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (Clinical Evaluation Exercise): A Preliminary Investigation. *Annals of Internal Medicine*. 1995;**123**:795-799.

[5]   Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*. 2007;**29**:855-871.

[6]   Kogan JR, Holmboe ES, Hauer KE. Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees. *JAMA: The Journal of the American Medical Association*. 2009;**302**:1316-1326.

[7]   Hatala R, Cook D, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Advances in Health Sciences Education*. 2015:1-27.

[8]   Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. Constructing a Validity Argument for the Mini-Clinical Evaluation Exercise: A Review of the Research. *Academic Medicine*. 2010;**85**:1453-1461.

[9]   Bok H, Teunissen P, Favier R, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education*. 2013;**13**:123.

[10]   Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014;**4**.

[11]   Alves de Lima A, Conde D, Costabel J, Corso J, Vleuten C. A laboratory study on the reliability estimations of the mini-CEX. *Advances in Health Sciences Education*. 2013;**18**:5-13.

[12]   Govaerts M, Schuwirth L, Van der Vleuten C, Muijtjens A. Workplace-based assessment: effects of rater expertise. *Advances in Health Sciences Education*. 2011;**16**:151-165.

[13]   Govaerts M. Workplace-Based Assessment and Assessment for Learning: Threats to Validity. *Journal of Graduate Medical Education*. 2015;**7**:265-267.

[14]   Yeates P, O'Neill P, Mann K, Eva KW. Effect of exposure to good vs poor medical trainee performance on attending physician ratings of subsequent performances.[Erratum appears in JAMA. 2013 Jan 16;309(3):237]. *JAMA*.**308**:2226-2232.

[15]   Yeates P, O'Neill P, Mann K, Eva K. Seeing the same thing differently. *Advances in Health Sciences Education*. 2013;**18**:325-341.

[16]   Gingerich A, van der Vleuten CPM, Eva KW, Regehr G. More Consensus Than Idiosyncrasy: Categorizing Social Judgments to Examine Variability in Mini-CEX Ratings. *Academic Medicine*. 2014;**89**:1510-1519.

[17]   Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*. 2012;**46**:201-215.

[18]   Menon S, Winston M, Sullivan G. Workplace based assessment: attitudes and perceptions among consultant trainers and comparison with those of trainees. *The Psychiatrist*. 2012;**36**:16-24.

[19]   Hassell A, Bullock A, Whitehouse A, Wood L, Jones P, Wall D. Effect of rating scales on scores given to junior doctors in multi-source feedback. *Postgraduate Medical Journal*. 2012;**88**:10-14.

[20]   Pelgrim EAM, Kramer AWM, Mokkink HGA, den Elsen L, Grol RPTM, Vleuten CPM. In-training assessment using direct observation of single-patient encounters: a literature review. *Advances in Health Sciences Education*. 2011;**16**:131-142.

[21]   Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. *BMC Medical Education*. 2014;**14**:76-76.

[22]   De Houwer J, Barnes-Holmes D, Moors A. What is learning? On the nature and merits of a functional definition of learning. *Psychon Bull Rev*. 2013;**20**:631-642.

[23]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Medical Education*. 2012;**46**:593-603.

[24]   Watling C. Cognition, culture, and credibility: deconstructing feedback in medical education. *Perspectives on Medical Education*. 2014;**3**:124-128.

[25]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education*. 2013;**47**:585-594.

[26]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Music lessons: revealing medicine's learning culture through a comparison with that of music. *Medical Education*. 2013;**47**:842-850.

[27]   Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgements. *Medical Education*. 2012;**46**:192-200.

[28]   Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments in Foundation Programme training: do trainees in difficulty use them differently? *Medical Education*. 2013;**47**:292-300.

[29]   Cleland J, Mackenzie RK, Ross S, Sinclair HK, Lee AJ. A remedial intervention linked to a formative assessment is effective in terms of improving student performance in subsequent degree examinations. *Medical Teacher*. 2010;**32**:e185-e190.

[30]   Barrett A, Galvin R, Steinert Y, et al. A BEME (Best Evidence in Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. *Systematic Reviews*. 2015;**4**:65.

[31]   Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA StatementThe PRISMA Statement. *Annals of Internal Medicine*. 2009;**151**:264-269.

[32]   Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. *Journal of the Medical Library Association*. 2006;**94**:421-e205.

[33]   Barr H, Freeth D, Hammick M, Koppel I, Reeves S. *Evaluations of interprofessional education; a United Kingdom review of health and social care*. London CAIPE/BERA; 2000.

[34]   Steinert Y, Mann K, Centeno A, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher*. 2006;**28**:497-526.

[35]   Buckley S, Coleman J, Davison I, et al. The educational effects of portfolios on undergraduate student learning: A Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. *Medical Teacher*. 2009;**31**:282-298.

[36]   Lo B. Informed Consent. *Ethical issues in clincial research: a practical guide* Lippincott Williams & Wilkins, Philadelphia, PA, 2012.

[37]   Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care* 2007;**19**:349-357.

[38]   Tong A, Palmer S, Craig JC, Strippoli GFM. A guide to reading and using systematic reviews of qualitative research. *Nephrology Dialysis Transplantation*. 2014.

[39]   CASP. Critical Appraisal Skills Programme 2013.

[40]   McHugh M. Interrater reliability: the kappa statistic. *Biochemica Medica*. 2012;**22**:276-282.

[41]   Yao DC, Wright SM. National survey of internal medicine residency program directors regarding problem residents. *JAMA*. 2000;**284**:1099-1104.

[42]   Brown DJ, Thompson RE, Bhatti NI. Assessment of operative competency in Otolaryngology residency: Survey of us Program Directors. *Laryngoscope*. 2008;**118**:1761-1764.

[43]   Hiemstra E, Kolkman W, Wolterbeek R, Trimbos B, Jansen FW. Value of an objective assessment tool in the operating room. *Canadian Journal of Surgery*. 2011;**54**:116-122.

[44]   Donato AA, Park YS, George DL, Schwartz A, Yudkowsky R. Validity and Feasibility of the Minicard Direct Observation Tool in 1 Training Program. *Journal of Graduate Medical Education*. 2015;**7**:225-229.

[45]   Young JQ, Lieu S, O'Sullivan P, Tong L. Development and initial testing of a structured clinical observation tool to assess pharmacotherapy competence. *Academic Psychiatry*. 2011;**35**:27-34.

[46]   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. *BJA: The British Journal of Anaesthesia*. 2009;**103**:524-530.

[47]   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**:744-749.

[48]   Whitehouse A, Hassell A, Bullock A, Wood L, Wall D. 360 degree assessment (multisource feedback) of UK trainee doctors: field testing of team assessment of behaviours (TAB). *Medical Teacher*. 2007;**29**:171-176.

[49]   Burford B, Illing J, Kergon C, Morrow G, Livingston M. User perceptions of multi-source feedback tools for junior doctors. *Medical Education*. 2010;**44**:165-176.

[50]   Archer JC, McAvoy P. Factors that might undermine the validity of patient and multi-source feedback. *Medical Education*. 2011;**45**:886-893.

[51]   Black D, Welch J. The under-performing trainee – concerns and challenges for medical educators. *The Clinical Teacher*. 2009;**6**:79-82.

[52]   Chipp E, Srinivasan K, Adil Abbas Khan M, Rayatt S. Incorporating multi-source feedback into a new clinically based revision course for the FRCS(Plast) exam. *Medical Teacher*. 2011;**33**:e263-e266.

[53]   Hesketh EA, Anderson F, Bagnall GM, et al. Using a 360 degrees diagnostic screening tool to provide an evidence trail of junior doctor performance throughout their first postgraduate year. *Medical Teacher*. 2005;**27**:219-233.

[54]   Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments of junior doctors: do scores predict training difficulties? *Medical Education*. 2011;**45**:1190-1198.

[55]   Vivekananda-Schmidt P, MacKillop L, Crossley J, Wade W. Do assessor comments on a multi-source feedback instrument provide learner-centred feedback? *Medical Education*. 2013;**47**:1080-1088.

[56]   Warm EJ, Schauer D, Revis B, Boex JR. Multisource feedback in the ambulatory setting. *J Grad Med Educ*. 2010;**2**:269-277.

[57]   Bullock AD, Hassell A, Markham WA, Wall DW, Whitehouse AB. How ratings vary by staff group in multi-source feedback assessment of junior doctors. *Medical Education*. 2009;**43**:516-520.

[58]   Ellul P, Moore A, Camilleri J, Grech V. An analysis of multisource feedback within the foundation programme. *Clin Teach*. 2012;**9**:290-294.

[59]   Wood L, Wall D, Bullock AD, et al. "Team observation": a six-year study of the development and use of multi-source feedback (360-degree assessment) in obstetrics and gynaecology training in the UK. United Kingdom, 2006.

[60]   Gingerich A, Regehr G, Eva KW. Rater-Based Assessments as Social Judgments: Rethinking the Etiology of Rater Errors. *Academic Medicine*. 2011;**86**:S1-S7.

[61]   Bullock A, Hassell A, Markham W, Wall D, Whitehouse A. How ratings vary by staff group in multi-source feedback assessment of junior doctors. *Med Educ*. 2009;**43**:516 - 520.

[62]   St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Medical Education*. 2015;**49**:548-550.

[63]   Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 2015;**49**:560-575.

[64]   Cook D, Zendejas B, Hamstra S, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education*. 2014;**19**:233-250.

CHAPTER 6

Discussion

Reconstructing the construct of formative assessment

This programme of research set out to answer a central research question: *what is the role and value of workplace-based assessment in learning in postgraduate medical education?* Evidence for the impact and value of workplace-based assessment (WBA) as a solution to the problem of limited observation-based feedback in postgraduate medical education remains elusive. The development, implementation and evaluation of WBA tools and their integration within training programme curricula has varied internationally; consequently, evidence for the intended learning value of this innovation has not been established. The aim of this series of research studies was to attempt to define and articulate the learning value of WBA to three key user groups: trainees (learners), trainers (teachers) and to the educational institutions (academy).

The research programme was designed to establish *what* was occurring in the implementation of WBA in Irish postgraduate medical training programmes, *how or why* this may have impacted on user perceptions of the learning value of the innovation and *how or why* WBA may be of value to trainers, trainees and training institutions in the ongoing challenge of detecting and remediating underperformance.

In this chapter I will discuss the main findings of the research studies and their implications for postgraduate medical education research and practice. The findings of the studies are of particular relevance to the outcome-based education movement. I propose that in order to better position WBA as a learning tool, its role and value must be clearly and consistently articulated within a planned programme of assessment and reconceptualised as a *practice* instead of a series of stand-alone tools and forms.

Main research findings and comparison to existing literature

Chapter 2 detailed the findings of the initial 'scene-setting' retrospective cohort study, designed to explore whether the context of the programme of research was consistent with internationally recognised patterns of WBA use. The study revealed patterns of completion in line with previously established literature; the ePortfolio records showed that the majority of WBA episodes were completed in the latter stages of the training year (after week 30) and 'clusters' of assessments were uploaded at one time. While trainee-related 'feedback' was provided on 44.9% of the recorded assessments, the quality of this feedback varied hugely, ranging from a single word (e.g. 'excellent') to complete sentences. Action plans or goal-oriented feedback were not included in any of the assessments and despite the mandatory requirements, only 63.8% of the trainees had documented evidence of engagement with WBA. These patterns were also in line with previous studies in which WBA was viewed as a 'tick box exercise'.[1]

These findings resonate with a recent programme of research in the Netherlands that explored the quality of feedback provided on assessment of residents in their first three years using the *CanMEDS* roles.[2, 3] Throughout this series of studies differences were apparent in the provision of feedback on certain roles, (most notably those of 'manager' and 'collaborator') and positive aspects of performance were noted more often than those aspects requiring attention.[2] In a second study,[3] the authors employed a critical discourse analysis framework to ascertain how supervisors constructed those roles and provided feedback; a number of interesting themes emerged including the emphasis on efficiency in service delivery and performance and the development of the trainee as a team leader. While the data available for analysis in our study could not be interrogated for such profiles, mainly due to the overwhelming absence of feedback, this Dutch study is of significance in informing ongoing development of outcome-based models of education and training, particularly relating to attempts to implement trainee-oriented feedback mechanisms, processes and practices.

One finding of the retrospective cohort study, however, was of particular significance. Forty (8.63%) of the WBA forms reviewed contained a rating of 'borderline' or 'below expectation'; 38 were Objective Structured Assessment of Surgical Skills (OSATS) forms and were from a single procedure-based specialty (obstetrics and gynaecology). At higher specialist training level, 10 of the 12 WBA forms containing these ratings were also OSATS specific to obstetrics and gynaecology.

This is the first study, to our knowledge, to report such findings. While it is acknowledged that trainers are reluctant to provide negative feedback for multiple and complex reasons,[4] it remains unclear in this study as to why trainers appeared to be more willing to document areas of concern at basic specialist level in single procedure-based disciplines, although not at higher specialist level. This is not necessarily evidence of poorer performance by trainees in this discipline; rather, the ceiling effect evident in all other specialties is potentially more concerning.

In interpreting this finding, it may be worth considering the current Irish context in which obstetrics and gynaecology training is now occurring. There have been a number of high profile adverse events and subsequent national reports on the quality of maternity services in Ireland over the last three years. As a result there are multiple new quality improvement initiatives and practices in place to improve safety and quality of care; for example, early warning scores have been introduced in response to a number of maternal and fetal deaths which may also influence trainers' willingness to articulate and report concerns about technical proficiency at an earlier stage.

The results of the cohort study informed the design of the subsequent phenomenological paper (**Chapter 3**), in which I set out to explore users' perceptions of the learning value of WBA and whether - or how - these perceptions had been shaped by experiences. While the context of this second study (the institution's implementation of WBA) was highly situated, the emergent data have broad resonance with a number of internationally-recognised issues. These include variations in understanding of the purpose of WBA and perceptions of the value of

its inclusion in training programme requirements. Similar to recent studies by Bok et al[5] and Rees et al,[6] trainers and trainees in this study understood the purpose of these assessments to be summative, with the aim of assuring competence. As evidenced in previous studies[7] *trainers* articulated mixed perceptions of the learning value of WBA – opinions in our study ranged from 'essential' to 'a complete waste of time'. Conversely, the perceptions of *trainees* were surprisingly positive – there was unanimous agreement that WBA was of potential value and should be an integral component of training requirements. The value however, was not specifically articulated in terms of learning, but rather in protecting training time and in being able to justifiably request feedback from their trainers. Interestingly trainers felt that they provided effective feedback without having to use WBA, where trainees felt that feedback was only provided when they sought it through the mechanism of WBA.

While this study provided some insights into the role of WBA in learning, we did not expect that it would be difficult for users to articulate their understanding or conceptualisation of learning. It appears that in postgraduate medical education, learning is perceived to be evidenced only by changes in practice.[8, 9] Although this is one valid outcome of the assessment-feedback process, the value of affirmation of good practice and determination of practice improvements over time are not yet clearly understood. In this study, trainers felt that trainees should learn from service provision, whereas trainees felt that service provision was of limited learning value compared to one-to-one 'training time'. The definition of learning within this unique educational context therefore requires further exploration if we are to better understand the formative role of WBA and its impact.

The final study in this research programme set out to understand whether workplace-based assessment is of value in the identification and/or remediation of underperformance among postgraduate medical trainees. *Chapter 4* detailed the development of the review protocol using the Best Evidence Medical Education (BEME) review methodology. *Chapter 5* described the review process and findings, the most significant of these being the diversity in research quality, heterogeneity of research methods and the surprising lack of comparative studies.

The findings of the review may be viewed under two separate categories. Firstly, while we were unable to make clear recommendations for the use of WBA to identify underperformance, the studies revealed such varied implementation and evaluation methods that it was also impossible to determine whether it is the content or nature of the individual tools, or the manner in which they have been implemented that may or may not be of use and value in this important aspect of postgraduate medical education and training.

The current literature supports the use of WBA as a formative assessment, designed to promote and inform learning and development, and appears to oppose the use of these tools and methods in making summative judgments of performance, particularly in the case of stand-alone assessment events. In the establishment of programmes of assessment, where WBA forms an integrated, continuous and longitudinal component of performance assessment, this may facilitate the determination of whether certain numbers of concern ratings, or types of concern ratings, are predictive of underperformance.

Secondly, it appears that the use of WBA in remediation of underperformance has not yet been thoroughly investigated. Of the 20 studies included in this review, one cross-sectional paper surveyed programme directors as to the availability of remediation mechanisms within their institutions. From the survey data (which were reliant on participant recall), it appeared that the availability of a remediation mechanism was associated with greater numbers of trainees identified as having performance-related issues. However, as noted in the review findings, the interpretation and transferability of their results were limited significantly by the study design and lack of triangulation with documentary evidence. The remediation outcomes were similarly under-described and it was not possible to extract the role (if any) of WBA in these processes.

It is difficult to compare the findings of this review with those of previous studies. While previous systematic reviews have concluded that WBA has limited impact on changing in performance[8-10] they did not explore the impact from a baseline of poor or underperformance. It is therefore unclear as to whether changes in practice

were required (remediation) or whether other learning effects (such as affirmation or consolidation of good practice) may have provided more impact in a cohort of performing doctors.

Implications of the research for postgraduate medical education

Assessment was commonly perceived to be the villain of workload…

Hayward, 2015[11]

The aim of this programme of research was to contribute to the discourse of WBA and to inform emerging and evolving postgraduate medical education assessment practices. The quote above very much reflects the starting point of this research programme, which was grounded in the problem of the acceptability, feasibility and limited implementation of WBA. The implications of the research are described firstly in terms of a local project to develop a model of outcome-based education for our institution and secondly, in the context of their implications for the wider global discussion on WBA.

Local implications

We are now fully aware of the barriers to the implementation of WBA in postgraduate medical training programmes. WBA is currently viewed by trainers and trainees as a series of individual and unconnected forms parachuted in to training programme curricula as single-event assessments of performance. However, the potential learning value associated with WBA as an ongoing and regular practice has stood out from our findings. To this end at our institution we plan to implement not just a series of mandatory requirements within our new model of outcome-based education, but to promote and support the development of observation-based feedback practices, empowering not only trainers but also trainees in the establishment of these practices.

There are also plans in place to better link the WBA tools, methods and practices to a new outcomes framework, using emerging research evidence for best practice in programmatic approaches to assessment.[5, 12-18] From our findings the absence of a defined framework in which the WBA could be situated and understood impacted on the implementation and subsequent experiences of WBA as a 'tick-box exercise' to be completed by the end of the training year. However, we have also noted in the work of Renting et al[2, 3] that that even with an established - and in their case, a well-recognised - framework, that this may not necessarily result in the delivery of the intended feedback. The communication of the purpose of any framework and consistent approaches to faculty development, along with trainee engagement in the design and implementation of the framework and associated programmes of assessment are a priority within this project.

International implications

1. Defining 'learning'

One of the key implications for both practice and research to emerge from these studies is the need to better define learning in postgraduate medical education and consequently to devise more appropriate methods to evaluate this learning. The most recent debate in the literature has suggested that a purely 'functional' or 'mechanistic' definition of learning as measurable changes in either behaviour or attitudes is limited; de Hower et al[19] have instead proposed that this definition be extended to one of 'ontogenic adaptation', that is, 'changes in the behaviour of an organism that result from regularities in the environment of the organism' (p633).

I propose that this definition however, is also limited by the use of the word 'changes'; in the unique context that is postgraduate medical education, Billett[20] suggests that workplace-learning, while occurring constantly, is also difficult to manage. He suggests that learners' reactions to learning events or activities for which they are not yet ready, and that are not within what Vygotstky termed the 'zone of proximal development', may be detrimental to learning. In our study,

learning was also constructed as affirmation of good practice along with identification of areas for improvement. By reconsidering the definition of learning within this context, the role and value of WBA may also lie in assisting in the identification of that 'zone of proximal development' for the individual trainee. For instance, the definition proposed by de Houwer et al could be reworded as 'changes *in or consolidation of* behaviour of the organism…'

2. Implications for feedback

The emphasis and value placed on feedback by trainees in our study are not new concepts; issues relating to feedback delivery[4] and acceptance[21-24] have been the subject of a large volume of international literature. The holy grail of effective feedback remains a long way off. In recent times with the evolution of newer validity arguments for WBA[25-28] and the potential acceptance of the value of subjective judgments that prioritise learning and development, we would concur with Hodges[29] that the regular practice of WBA, in which a culture of assessment and feedback is normalised,[24] is of significant individual value and that validity in this sense should not be dismissed.

This proposal will create challenges for regulators and training bodies. Over the last 15 years the literature on WBA has in many ways responded to the need to improve accountability on the part of such bodies and organisations, with WBA seen as a potential solution to both assessment of competence and acceding to trainee demands for better feedback. But it appears that WBA cannot effectively fulfil both sets of demands and that other assessment methods are required if the academies wish to provide standardised assessment and that in the longer term, the evolution of programmatic approaches to assessment[16], while they may be more onerous and work-intensive to implement, may be the better solution to ensuring competence.

The impact of rater variability on the credibility and validity of WBA has provided a second large area of study,[30-33] but this work is predicated on the assumption that WBA must be viewed as a summative judgment and therefore one that is seen to

require levels of objectivity and distance from the trainee. A recent study by Hawkins et al[34] suggests, somewhat tentatively, that this shift away from a summative towards a more formative purpose, with a consequent shift in the way that learning and the value of WBA are perceived, may have started to occur as WBA becomes embedded in the surgical training culture. The recent change in the conversation on assessment-*for*-learning to one of assessment-*as*-learning may also eventually impact on the way in which assessment is viewed,[11] particularly in the context of longitudinal integration within programmes of assessment.

3. Reconstructing WBA

As I have alluded to throughout this research thesis, the implementation of WBA internationally has utilised a series of individual tools, each designed to assess trainee competence in specific skills or clinical tasks. The most obvious difficulty with WBA in this form is that in order to provide 'reliable' judgments of performance the workload of assessment and form-filling for trainers has become overwhelming. My research proposes that trainees see potential value in WBA as a structure for observation-based feedback, but that it is the time between trainer and trainee, in which they can demonstrate their competence to trainers and receive feedback that is of most value, a constructionist viewpoint which departs from the positivist implementation strategies commonly employed.

Recent evidence from the UK[6] has also suggested that changing the title of the assessments (from WBA to 'supervised learning events' or SLEs) is ineffective without communication of the formative nature of the events. In order to maximise on this potential value and engagement by trainees, we need to better reconceptualise WBA. To this end it is worth considering Bennett's[35] proposed definition of formative assessment as 'a thoughtful integration of process and purposefully designed methodology or instrumentation'; in this case, WBA could potentially be defined as a 'practice'. As with all practices, normalising the assessment-feedback loop will require skills and competencies of both trainers and trainees to be maximally effective and as programmatic approaches to assessment

continue to develop, the continuous assessment-feedback cycles may assist in this normalisation process.

4. Long-term value in normalising assessment and feedback

Assessment of performance is a stressful event, mainly for trainees or learners, but also for those tasked with the responsibility of judging performance and assuring competence. In low-stakes settings, such as the naturalistic work settings in which WBA occurs, this appears to also be the case. Normalising the practice of WBA as proposed above may assist in limiting this stress in postgraduate medical education, contingent on a number of conditions being understood and accepted by all participants. These include:

- Consistent communication of the use and purpose of WBA – is it summative or formative, high-stakes or low-stakes?
- Consistent approaches to the documentation of WBA and rating scales including the merits and demerits of numerical scores or narrative feedback
- Longitudinal implementation of these practices. As the potential revalidation and maintenance-of-competence processes emerge over the next couple of years, all doctors in practice will be required to submit evidence of peer review and the ability to seek and use feedback.[36] Normalising the practice of observation, assessment and feedback at training stages may assist in the implementation of these processes.

Implications for future research

1. Issues of authenticity of WBA

The implementation of WBA was proposed as formative assessment in an 'authentic' work-based setting. I would argue that the WBA within the current implementation frameworks is far from authentic. While the setting may indeed be authentic, in that WBA is carried out in the trainee's usual clinical practice setting and with real (as opposed to simulated or standardised) patients, I discovered that assessments, particularly in the case of the mini-CEX, are more often than not 'set up'. The trainer and trainee agree to complete a WBA, a patient's consent is obtained and the trainer is obliged to stand back and observe while the trainee leads on the patient encounter. The trainee is therefore in the position of being assessed, while the trainer assesses and potentially stands behind the trainee with pen in hand to complete the form. In the usual setting and practice, the trainee often carries out a physical examination or performs a procedure without having to look for consent to do so as an assessment event. Already, the context has changed. In the qualitative study described in Chapter 3, trainees alluded to this artificial context as a contributory factor to their perception of the limited learning value of WBA but this will require further study to establish whether these patterns are under-recognised as a contributory factor or if this is purely a finding relate to the local context.

This contradiction reflects that described by Hodges[37] in the context of OSCEs and whether or not assessing that the doctor *can* perform the procedure or task is evidence that this is *actually* how they do it when un-observed. In his paper, Hodges invokes Goffman's theory of the 'the presentation of the self'.[38] Goffman contended, using the metaphor of the stage, that all 'actors' have two personas - a back-stage and a front-stage persona and that these personas demonstrate differing characteristics depending on the location of the performance, the actor's perceptions of the expectations of the audience, the demands of the role and the interpretation of that role. He compared this to the OSCE setting in which the situation is simulated to mimic 'real life', be that a real clinical problem or simulated

procedure; but by the nature of the examination context, the presence (either actually in the room or via video) of examiners and the stakes involved, OSCEs cannot be deemed to be authentic. The learner (or actor), particularly, for example, in the case of a station where the trainee or learner must examine or communicate a diagnosis to a patient, is performing for two people – the 'patient' and the examiner. The learner makes a decision therefore, as to which persona they need to engage to complete the performance.

Hodges' paper could be rewritten for WBA. In its current guise, WBA often involves a 'staging' of an assessment event which, even with the lowest stakes in mind, automatically becomes one in which the trainee determines the persona they will engage for the purposes of that event. Goffman also suggests that in undertaking an established role (i.e. that of the medical trainee), the actor usually references already-created 'fronts' or personas based on previous experience (p37). Whether or not WBA performance therefore reflects actual daily practice has not been fully established and requires further examination. Conversely, multi-source feedback tools, including the Team Assessment of Behaviour were developed to obtain a sense of the trainee's performance over time. By including a wide variety of assessors who provide a judgment of performance based on more than a single event, the assessment may provide a better picture of the learner's consistency in performance and situate the assessor as less of a critic and more as a director aiming to improve that performance.

2. To be (formative) or not to be (summative); is that the question?

To continue our dramaturgical metaphor, it appears that the 'stage' of WBA needs to be consistently communicated. If this is to be a continuous, low-stakes assessment practice in which learning is prioritised, then the associated actors' 'scripts' require changes and flexibility to respond to individual trainee learning and development needs. Given the widely varied implementation methods of WBA across the world, it appears that it is timely to consider a consensus on the purpose of WBA to all its users.

It also appears that the roles of the players within the WBA drama can then be redefined. From our qualitative study, issues of power and identity emerged during the analysis that we had not previously considered. Using WBA to request – and justify the request for – protected training time is an interesting narrative and commentary on the current role of trainees within a health service, particularly within our national context. Emerging research from Canada and the United States[39] has already begun to explore the links between identity development and the pressures of service provision and it appears that in our current context, trainees are required to assume multiple identities – as learner, service provider and subject matter expert. Further research could explore the links between formative assessment practices and trainees' perceptions of their identity.

3. Medical education and formative assessment

There is a vast, growing body of formative assessment literature across educational fields and contexts including emerging theories of formative assessment that were considered in the initial proposal for this thesis.[35, 40, 41] While this literature is of potential value to the development of a theory of WBA, the context of medical education - and specifically that of postgraduate training – is unique.[20] Even among postgraduate medical education models, the transferability of research findings can be troublesome; as I have discussed throughout this thesis, implementing WBA requirements to meet numbers required for reliability has not been feasible within our context. Research evidence emerging from large data sets and from training jurisdictions or cultures that place an emphasis on training over service provision - or at the very least attempt to balance the two - cannot easily be generalised to a context in which service provision trumps training. Research efforts therefore need to consider the multiple and varied contexts of postgraduate medical education and the place, value and role of formative assessment throughout all settings.

4. Conditions required to improve the learning value of WBA among underperforming trainees – maximising the opportunity provided by programmatic approaches to assessment

In this thesis, I sought to understand whether WBA tools and methods could improve learning from the baseline of underperformance ~~in the context of remediation;~~ our BEME review revealed that this remains a largely understood area of learning and development. The emergence and increasing adoption of programmatic approaches to assessment may provide a framework in which WBA may be used to support learning among this specific cohort.

Developed over the last ten years, the theory of programmatic assessment proposes the design of assessment structures that reduce reliance on single-event, high-stakes assessments to make progress or competence decisions, focused on assessment *of* learning; rather, the inclusion of multiple, regular and low-stakes assessments, focused on providing a learning value, are then taken together to form an overall illustration of the learner's progress. Progression or competence decisions are made based on the 'sum of the parts' rather than individual data points.[15, 17]

While the implementation of programmes of assessment has proved challenging[5, 18] and requires large-scale institutional and cultural change[42] early empirical evidence suggests that in general:

1. Students value the learning-centred continuous low-stakes assessments [43, 18]

2. Programmatic approaches appear to assist in early identification of underperformance and reduce the incidence of 'failure to fail' when progression decisions are taken by a panel of assessors who can look at the entire picture of trainee performance over time[44]

Although features of programmatic assessment *implementation* e.g. delayed feedback, may also inhibit learning[18] current guidelines suggest that the appointment of a mentor within this process is key in facilitating the interpretation

of feedback and multiple low-stakes assessments among learners.[18, 43] While evaluation of these approaches is ongoing, I also propose that key to the success of this medical innovation is an understanding of how, or why, this approach may better facilitate learning among trainees than the current process of stand-alone WBA. Although recent studies have identified graduate-entry students' perceptions of the elements of programmatic assessment that may support or inhibit learning, this has not been explored among postgraduate cohorts, for whom the majority of the learning and assessment experiences are situated within the workplace.

Where postgraduate programmes of assessment have been implemented and established, it would be of interest to explore how trainees construct learning within this framework. Using a constructivist grounded theory (CGT) approach, underpinned by established and emerging theories of workplace learning (including the experience, trajectories and reifications (ETR) framework we used in this programme of research[45]). By understanding the learning impact of programmatic assessment approaches among the general trainee cohort, we may then be better able to research specific remediation and learning support programmes for those trainees who are, or are at risk of, underperforming. Indeed this study design may allow for a retrospective analysis of any differences in how performing and underperforming trainees construct learning within this framework.

5. Evolving methods in medical education research

The relatively recent emergence and growing acceptability of qualitative approaches in medical education research has facilitated significantly different conversations that have impacted on the practice of medical education over the last twenty years. Methodological evolution is continuing now, for example, in the change in concepts of validity that are now being explored, particularly in the case of WBA. Academics and researchers no longer ask 'does that work', but rather 'is it effective for this group or individual, in this context and how?' The acceptability of newer methods and approaches to research is also reflected in the design of this thesis – developed as a series of studies instead of a traditional larger study – which

allowed me to explore my research question from multiple angles and perspectives. By using varied study methods, I was able to understand the benefits and limitations of each, along with contradictions and similarities in the findings. However, I did not consider this a 'mixed methods' thesis in its traditional sense; in this study I asked different questions with the aim of understanding the value of WBA from these multiple perspectives; I did not use two sets of data to approach the same research question. However, the philosophical debate on what exactly constitutes mixed methods research appears to be continuing.[46, 47] Burke-Johnson and Onwuegbuzie[46] have recently called for a 'pragmatic approach' to mixed methods research in education and in this thesis, I would describe my approach as a 360° perspective on the value of WBA in learning in postgraduate medical education.

Limitations of the research

In Chapter 1 I sought to articulate my own research orientation and paradigm preferences that guided the development of this programme of research, which explored the role and value of WBA in learning from multiple perspectives and theoretical frameworks. While my own preferences towards constructivist approaches to research guided the initial study designs, the studies were also guided by the need to answer our questions in a way that would be accessible to information users from all paradigmatic orientations.[48] At each stage of the process I have attempted to ensure that these decisions were transparent and that the research methods employed were thorough and robust. However I acknowledge that research uncertainty will always exist and in this section, I aim to outline how I managed that uncertainty and how these findings could potentially be interpreted differently within different research contexts.[49, 50]

I explored my research question from three perspectives using three specific methodological approaches and theoretical frameworks. The retrospective cohort study was quantitative in its approach to data extraction and centred around a framework of 'good feedback practice'[51] but was limited to *what was written* on

the assessment forms. The interpretation of this study could be further enhanced in the future by considering what was *not* written in these WBA records. By employing an ethnographic or other immersive approach to the research, we could determine whether those feedback principles are indeed adhered to at the time of the WBA event through verbal feedback, and whether trainers feel that having provided that feedback verbally, the need to record it in logbook format is unnecessary and duplicative.

The phenomenological study was designed to explore trainers' and trainees' experiences of WBA with the aim of understanding how those experiences may have shaped perceptions of learning value. Grounded in an emerging theoretical framework of experience, trajectories and reifications (ETR),[45] we made an assumption from our initial study that WBA was now a core feature of workplace learning. During the study however, it emerged that this was not yet the case, limiting our ability to interpret the findings at all three levels of framework.

The emergence of two other issues also now requires further exploration; firstly, finding an acceptable and consistent definition - and by extension, new evaluation methods - of 'learning' in the unique educational context that is postgraduate medical education and secondly, the 'story we did not tell'[49] – one of issues of power, agency and professional identity that may form larger barriers to the effective implementation of WBA than we had anticipated. To this end, a critical discourse analysis approach to the study may uncover a very different understanding of the role of experiences in shaping perceptions.

The BEME review methodology was chosen to determine the value of WBA in identifying and/or remediating underperformance in an attempt to explore the research question from a wider international context. It was clear from the first two studies that the information available in our own local context would not be extensive enough to be able to make generalisable and transferable findings using observational or intervention-based studies for this specific review question. While the review methodology afforded a robust and logical approach to the question and included a number of qualitative studies, the findings are also limited by this

quantitative approach; given the individual nature of this formative assessment innovation it may be that the learning value to trainees who are underperforming may be more subtle and that the educational outcomes may be less tangible than this BEME review could discern.

Conclusion

Three years ago I set out to ask whether or not WBA has been realised as a formative educational tool in the way that it was originally intended. It is now evident that this has not yet happened and that WBA is not routinely used – or perceived to be of use – purely for learning purposes. However, it is also evident that this is not solely related to the tools involved or the implementation processes employed; rather the barriers to this realisation are complex, multifaceted and have their genesis in the fact that different user groups want different things from WBA. Implementation, evaluation and interpretation of the value of WBA have therefore also varied widely and there are many understandings and perceptions of the purpose of WBA around the world.

This programme of research has uncovered a number of findings which will contribute to the international conversation on WBA. Firstly, trainees perceive a potential learning value in WBA and in an era of overstretched health services, the training value that WBA affords them in the guise of a justifiable way to ask for feedback and time with their trainers is equally as valued. Secondly, WBA has not been shown to be able to identify underperformance as a stand-alone method of assessment. Instead it may assist, as part of a longitudinal programme of assessment, in identifying or remediating underperformance but the research methods used to determine this value will require some new and varied approaches. Thirdly, and in line with much international research, it is the *practice* of providing observation-based feedback that is of value to trainees, not the individual tools used to provide that structure.

In a time of unprecedented change in medical education practice and research, WBA is important to all stakeholders, but in order to be of value to trainers, trainees and ultimately to patients, the medical education community needs to establish the most effective position of WBA within programmes of assessment.

References

[1]   Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise? *Medical Teacher*. 2011;**33**:919-927.

[2]   Renting N, Gans ROB, Borleffs JCC, Van Der Wal MA, Jaarsma ADC, Cohen-Schotanus J. A feedback system in residency to evaluate CanMEDS roles and provide high-quality feedback: Exploring its application. *Medical Teacher*. 2015:1-8.

[3]   Renting N, Dornan T, Gans ROB, Borleffs JCC, Cohen-Schotanus J, Jaarsma ADC. What supervisors say in their feedback: construction of CanMEDS roles in workplace settings. *Advances in Health Sciences Education*. 2015:1-13.

[4]   Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*. 2012;**46**:201-215.

[5]   Bok H, Teunissen P, Favier R, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education*. 2013;**13**:123.

[6]   Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monrouxe LV. Supervised learning events in the Foundation Programme: a UK-wide narrative interview study. *BMJ Open*. 2014;**4**.

[7]   Menon S, Winston M, Sullivan G. Workplace based assessment: attitudes and perceptions among consultant trainers and comparison with those of trainees. *The Psychiatrist*. 2012;**36**:16-24.

[8]   Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;**341**.

[9]   Overeem K, Wollersheim H, Driessen E, et al. Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study. *Medical Education*. 2009;**43**:874-882.

[10]   Saedon H, Salleh S, Balakrishnan A, Imray C, Saedon M. The role of feedback in improving the effectiveness of workplace based assessments: a systematic review. *BMC Medical Education*. 2012;**12**:25.

[11]   Hayward L. Assessment is learning: the preposition vanishes. *Assessment in Education: Principles, Policy & Practice*. 2015;**22**:27-43.

[12]   ten Cate O. The false dichotomy of quality and quantity in the discourse around assessment in competency-based education. *Advances in Health Sciences Education*. 2014:1-4.

[13]   Touchie C, ten Cate O. The promise, perils, problems and progress of competency-based medical education. *Medical Education*. 2016;**50**:93-100.

[14]   van der Vleuten C, Verhoeven B. In-training assessment developments in postgraduate education in Europe. *ANZ Journal of Surgery*.**83**:454-459.

[15]   Dijkstra J, Van der Vleuten C, Schuwirth L. A new framework for designing programmes of assessment. *Adv Health Sci Educ*. 2009;**15**:379 - 393.

[16]   Schuwirth L, Van der Vleuten C. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;**33**:478 - 485.

[17]   van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Medical Teacher*. 2012;**34**:205-214.

[18]   Heeneman S, Oudkerk Pool A, Schuwirth LWT, van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Medical Education*. 2015;**49**:487-498.

[19]   De Houwer J, Barnes-Holmes D, Moors A. What is learning? On the nature and merits of a functional definition of learning. *Psychon Bull Rev*. 2013;**20**:631-642.

[20]   Billett S. Learning through health care work: premises, contributions and practices. *Medical Education*. 2016;**50**:124-131.

[21]   Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgements. *Medical Education*. 2012;**46**:192-200.

[22]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Medical Education*. 2012;**46**:593-603.

[23]   Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Medical Education*. 2013;**47**:585-594.

[24]   Watling C. Cognition, culture, and credibility: deconstructing feedback in medical education. *Perspectives on Medical Education*. 2014;**3**:124-128.

[25]   Cook D, Zendejas B, Hamstra S, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education*. 2014;**19**:233-250.

[26]   Hatala R, Cook D, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Advances in Health Sciences Education*. 2015:1-27.

[27]   Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. 2015;**49**:560-575.

[28]   St-Onge C, Young M. Evolving conceptualisations of validity: impact on the process and outcome of assessment. *Medical Education*. 2015;**49**:548-550.

[29]   Hodges B. Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical Teacher*. 2013;**35**:564-568.

[30]   Yeates P, O'Neill P, Mann K, Eva K. Seeing the same thing differently. *Advances in Health Sciences Education*. 2013;**18**:325-341.

[31]   Gingerich A, Regehr G, Eva KW. Rater-Based Assessments as Social Judgments: Rethinking the Etiology of Rater Errors. *Academic Medicine*. 2011;**86**:S1-S7.

[32]   Gingerich A, van der Vleuten CPM, Eva KW, Regehr G. More Consensus Than Idiosyncrasy: Categorizing Social Judgments to Examine Variability in Mini-CEX Ratings. *Academic Medicine*. 2014;**89**:1510-1519.

[33]   Govaerts M. Workplace-Based Assessment and Assessment for Learning: Threats to Validity. *Journal of Graduate Medical Education*. 2015;**7**:265-267.

[34]   Hawkins S, Osborne A, Egbeare D, Williamson J, Lambert A. A positive change in surgical attitudes towards online workplace-based assessment. *Bulletin of the Royal College of Surgeons of England*. 2014;**96**:e6-e9.

[35]   Bennett RE. Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*. 2011;**18**:5-25.

[36]   Norcini J. Understanding learning in the workplace for those who practise: we can't wait another 50 years. *Medical Education*. 2016;**50**:18-20.

[37]   Hodges B. OSCE! Variations on a theme by Harden. *Medical Education*. 2003;**37**:1134-1140.

[38]   Goffman E. *The Presentation of Self in Everyday Life*. London, New York Double Day; 1959.

[39]   Kesselheim JC, Sun P, Woolf AD, London WB, Boyer D. Balancing Education and Service in Graduate Medical Education: Data From Pediatric Trainees and Program Directors. *Academic Medicine*. 2014;**89**:652-657.

[40]   Clark I. Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*. 2012;**24**:205-249.

[41]   Black P. Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*. 2015;**22**:161-177.

[42]   Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Medical Education*. 2016;**50**:702-708

[43]   Driessen EW, van Tartwijk J, Govaerts M, Teunissen P, van der Veluten CPM. The use of programmatic assessment in the clinical workplace: A Maastricht case report. *Medical Teacher*. 2012;**34**:226-231

[44]   Wilkinson TJ, Tweed MJ, Egan TG, Ali AN, McKenzie JM, Moore M, Rudland JR. Joining the dots: Conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Medical Education*. 2011; **11**:29

[45]   Teunissen P. Experience, trajectories, and reifications: an emerging framework of practice-based learning in healthcare workplaces. *Advances in Health Sciences Education*. 2014:1-14.

[46]   Johnson RB, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*. 2004;**33**:14-26.

[47]   Hesse-Biber S. Mixed Methods Research: The "Thing-ness" Problem. *Qualitative Health Research*. 2015;**25**:775-788.

[48]   Albert M, Hodges B, Regehr G. Research in Medical Education: Balancing Service and Science*. *Advances in Health Sciences Education*. 2006;**12**:103-115.

[49]   Helmich E, Boerebach BCM, Arah OA, Lingard L. Beyond limitations: Improving how we handle uncertainty in health professions education research. *Medical Teacher*. 2015;**37**:1043-1050.

[50]   Lingard L. The art of limitations. *Perspectives on Medical Education*. 2015;**4**:136-137.

[51]   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**:199-218.

APPENDICES AND FIGURES

Appendix 1   RCPI Mini-Clinical Evaluation Exercise

FORM 023 - GUIDELINES FOR MINI CLINICAL ASSESSMENT EXERCISE (Mini-CEX)

A mini-CEX assesses core skills that can be demonstrated during a routine encounter with a patient. It is based on a 15 minute observation of a single interaction. It may not be possible to explore all elements listed on every occasion. However, a score should always be given for the trainee's "overall competence". And though knowledge is not specifically tested, it is understood that it may well affect the standard of the performance delivered. **Scoring should reflect the performance of the trainee against that expected at their stage of training and experience.**

Notes for the Assessor:

(please read)

1. The assessment should take place in the usual place of work (*in-patient, clinic, office or department*) where the assessor must directly **observe** the trainee's performance.
2. Some indication of the nature and complexity of the problem faced should be given.
3. A range of descriptors are given to evaluate the trainee's performance at the current stage of training. In scoring the performance the overall difficulty of the case can be taken into account.
4. A space is provided for any comment **the assessor** wishes to add. This is particularly necessary if any mark has been given that is below expectations.
5. The assessor must give feedback to the trainee on the performance assessment just completed, dealing fully with any deficiency identified. Both the trainer and the trainee sign the form to confirm that this has taken place, and space is provided here for any additional comments from the trainee and trainer.
6. The fully completed form is handed to the trainee. It is their responsibility to keep a copy for their portfolio

Reflective Practice: Shows analytical, constructive approach to case, willingness to learn; acknowledges and prepared to consider other management options; aware of change, possible advances, when to seek help.

Competencies to be Assessed:

Consideration/Professionalism: Recognises/accepts patient's rights (to consent, confidentiality, and information). Establishes trust, shows professional approach

Communication: Informs, explains, and advises using appropriate language. Obtains consent, enlists patient's co-operation.

Interviewing Skills: "Active" listening facilitating relevance; effectively using questions, responding to non-verbal clues.

Examination Skills: Prepares patient, minimises discomfort/unease. Proceeds logically, efficiently, thoroughly, completely.

Judgment: Correctly identifies/lists problems, prioritises actions in realistic and timely schedule.

Please note: *While it is not always possible to assess all the elements listed at each encounter the importance of accurate, legible records of the facts of a case cannot be overlooked. Likewise, willingness to review one's actions constructively and to learn from this experience are valuable assets. Every effort should be made to include an evaluation of communication skills, due consideration given for the patient and a professional approach shown as well as evaluating the trainee's overall capability.*

MINI CLINICAL ASSESSMENT EXERCISE (Mini-CEX)

Trainee's Name:							

RTN:		Training Year:	1	2	3	4	5

Date of Assessment:		Location of Assessment:	

Complexity:	Low		Average		High	

Problem List:					
1.		2.		3.	
4.		5.		6.	

Discussed at:	Clinic	Ward	Office	Meeting	Other:

Mark **one** of the boxes for each of the categories below from. Marks that are well below expectations, below expectations and borderline for stage of training are considered unsatisfactory and if any of these marks are given the assessor **must** explain in the space provided below for his/her comments. **The marks given should fairly represent the trainee's performance in this evaluation** as expected for their stage of training.

	Well Below Expectation	Below Expectation	Borderline	Meets Expectation	Above Expectation	Well Above Expectation
Interviewing Skills ☐ Not Assessed						
Examination Skills ☐ Not Assessed						
Consideration/ Professionalism ☐ Not Assessed						
Communication ☐ Not Assessed						
Judgment ☐ Not Assessed						
Overall Competence						

Comments:		
Signed by Trainee:		**Date:**
Signed by Assessor:		**Date:**

FORM 021 - GUIDELINES FOR DIRECT OBSERVATION OF PROCEDURAL SKILLS (DOPS) ASSESSMENTS - CLINICAL

Direct observations of practical skills assess the capabilities of a trainee while he/she performs a procedure. The DOPS is a structured assessment of actual performance. The basic competencies required for most procedures can be assessed using the list of CORE SKILLS below; specific skills relevant to a particular procedure may be added.

Notes for the Assessor:

(Please read)

1. Please ensure the patient (client) is aware that a DOPS is taking place, and that the assessor has been properly introduced.

2. The assessor must be familiar with performing the procedure under assessment, also competent, in case the trainee requires assistance to complete.

3. The assessment should be made under appropriate conditions (e.g. with all equipment, personnel, necessary to support) and in a suitable environment.

4. The task to be undertaken should be of a degree of difficulty appropriate to the operator's capabilities.

5. A range of descriptors are given to evaluate the trainee's performance at the current stage of training. In scoring the performance the overall difficulty of the case can be taken into account.

6. A space is provided for any comment **the assessor** wishes to add. This is particularly necessary if any mark has been given that is below expectations.

7. The assessor **must** also give feedback to the trainee on the performance assessment just completed, dealing fully with any deficiency identified. Both the **trainer** and the trainee sign the form to confirm that this has taken place, and space is provided here for any additional comments from the trainee and the trainer.

8. The **fully completed** form is handed to the trainee. It is their responsibility to keep a copy for their portfolio.

Competencies to be Assessed:

Understanding of Procedure: Relevant anatomy; purpose, indications, contra-indications; outcomes, risks, complications; choice of methods available.

Consideration for the Patient: Gives reassurance, minimises discomfort, explains procedure fully; confirms informed consent obtained.

Preparation: First re-checks all relevant details correct. Safety check; instrumentation, equipment (drugs); positioning; cleansing/aseptic technique; sedation, analgesia, anaesthesia confirmed.

Professional/Technical Ability: Dexterity, accuracy, efficiency; obtains, interprets diagnostic material/information; informs, directs staff courteously; recognises own limitations; manages risk.

Post-Procedure: Completes documentation; regulates recovery phase, observations; anticipates/deals with complications. Informs/counsels patient/relatives.

Overall Ability to Perform Procedure: Ability to complete/undertake procedure; technical abilities as demonstrated; appropriately confident, team/ leadership skills.

DIRECT OBSERVATION OF PROCEDURAL SKILLS (DOPS)

Trainee's Name:							

RTN:	Training Year:	1	2	3	4	5
Date of Assessment:	Location of Assessment:					

Procedure Undertaken:

Complexity:	Low	Average	High	New Case	Review

Mark **one** of the boxes for each of the categories below from. Marks that are well below expectations, below expectations and borderline for stage of training are considered unsatisfactory and if any of these marks are given the assessor **must** explain in the space provided below for his/her comments. **The marks given should fairly represent the trainee's performance in this evaluation** as expected for their stage of training.

	Well Below Expectation	Below Expectation	Borderline	Meets Expectation	Above Expectation	Well Above Expectation
Selection of Procedure Relevance to patient's problems. □ Not Assessed						
Obtained Informed Consent □ Not Assessed						
Preparation □ Not Assessed						
Technical Ability □ Not Assessed						
Awareness & Management of Complications □ Not Assessed						
Post Procedural Advice □ Not Assessed						
Professionalism Consideration for the patient. □ Not Assessed						
Procedure notes Accuracy & detail. □ Not Assessed						
Overall competence						

Comments:		
Signed by Trainee:		**Date:**
Signed by Assessor:		**Date:**

FORM 020 - GUIDELINES FOR ASSESSMENT DURING CASE-BASED DISCUSSION (CBD)

CBD is used to evaluate core skills that can be demonstrated during an interactive discussion based on a single case in which the trainee has been actively involved. The case for discussion can either be selected by the trainee or chosen by the assessor. The assessment will be based on oral discussion and written information available. It includes a **bi-lateral** (trainee's and trainer's) critical appraisal of the reasoning and judgements made, and of the management of the case. It may not be possible to explore all the elements listed on every occasion. However, a score should always be given for the trainee's "overall competence" and although knowledge is not specifically tested, it is understood that it may well affect the standard of the performance delivered. **Scoring should reflect the performance of the SHO against that expected at their stage of training and experience.**

Notes for the Assessor:

(please read)

1. The presentation should take place in a suitable environment, with due consideration given to the patient's (client's) sensitivities, to confidentiality (e.g. in any ward or clinical setting; an office, side- or seminar-room may be found convenient).

2. Some indication of the nature and complexity of the problem faced should be given.

3. A range of descriptors are given to evaluate the trainee's performance at the current stage of training. In scoring the performance the overall difficulty of the case can be taken into account.

4. A space is provided for any comment **the assessor** wishes to add. This is particularly necessary if any mark has been given that is below expectations.

5. The assessor **must** also give feedback to the trainee on the performance assessment just completed, dealing fully with any deficiency identified. Both the **trainer** and the trainee sign the form to confirm that this has taken place, and space is provided here for any additional comments from the trainee and trainer.

6. The **fully completed** form is handed to the trainee. It is their responsibility to keep it in their portfolio.

Competencies to be Assessed:

Problem Definition: All relevant facts established, from current/previous history, investigations, interventions; reports, correspondence reviewed.

Record Keeping: Legible, tidy, legally defensible records seen.

Reasoning: Appropriately selected, sequenced investigations/procedures planned. Evidence-based, logical judgements made; (differential) diagnosis established; action plan made with realistic goals.

Case Management: Effective, safe (responsible) prescribing; aware of protocols/guidelines, best practice; monitoring progress, handling complications/mistakes; timely, appropriate referrals, case closure.

Reflective Practice: Shows analytical, constructive approach to case, willingness to learn; acknowledges and prepared to consider other management options; aware of change, possible advances, when to seek help.

Please Note: While it is not always possible to assess all the elements listed at each encounter the importance of accurate, legible records of the facts of a case cannot be overlooked. Likewise, willingness to review one's actions constructively and to learn from this experience are valuable assets.

CASE-BASED DISCUSSION (CBD)

Trainee's Name:						
RTN:		**Training Year:**	**1**	**2**	**3**	**4**	**5**

RTN:	Training Year:	1	2	3	4	5

Date of Assessment:	Location of Assessment:

Case Discussed:

Complexity:	Low	Average	High

Problem List:

1.	2.	3.
4.	5.	6.

Discussed at:	Clinic	Ward	Office	Meeting	Other:

Mark **one** of the boxes for each of the categories below from. Marks that are well below expectations, below expectations and borderline for stage of training are considered unsatisfactory and if any of these marks are given the assessor **must** explain in the space provided below for his/her comments. **The marks given should fairly represent the trainee's performance in this evaluation** as expected for their stage of training.

	Well Below Expectation	Below Expectation	Borderline	Meets Expectation	Above Expectation	Well Above Expectation
Problem Definition □ Not Assessed						
Record Keeping □ Not Assessed						
Reasoning □ Not Assessed						
Case Management □ Not Assessed						
Reflective Practice □ Not Assessed						
Overall Competence						

Comments:		
Signed by Trainee:		**Date:**
Signed by Assessor:		**Date:**

FORM 36 GUIDELINES FOR ASSESSMENT DURING OSATS (OPERATIVE LAPAROSCOPY)

There are a small number of procedures which are so fundamental to the practice of obstetrics and gynaecology that an objective assessment tool has been developed by Royal College Of Obstetricians & Gynaecologists, UK to aid the assessment process. The OSATS is a validated assessment tool to assess technical competency in a particular technique. The curriculum indicates those skills which need to be assessed with OSATS. The forms are included within the relevant module in the logbook. The OSATS should be used to help you and your trainer assess when you are ready to move on to independent practice for a procedure and when you are ready to be signed off for independent practice. The same OSATS may be used to assess increasing levels of complexity for any particular procedure.

Ten OSATS have been developed to assess those procedures that are fundamental to the practice in obstetrics and gynaecology. These are:

- fetal blood sampling
- diagnostic hysteroscopy
- diagnostic laparoscopy
- opening and closing the abdomen
- uterine evacuation
- perineal repair
- caesarean section
- operative vaginal delivery
- operative laparoscopy
- manual removal of the placenta.

Before the competences can be signed off in the logbook, each OSATS must have been successfully completed (that is, every box ticked for independent practice). When the trainee feels ready to undertake the relevant OSATS, they will meet with your clinical trainer who will assess the procedure and complete the OSATS form. A record of the date that each OSATS is signed off should be entered in the relevant section of the logbook module. At least two different assessors need to be involved for this process. The same assessor must not be used for all OSATS and a consultant must do at least one OSATS. Once signed up for independent practice it is recommended that, to demonstrate continued competency in an area, an annual OSATS assessment is performed.

Before undertaking an OSATS assessment the trainee must be able to perform the procedure competently under direct supervision. They will be required to demonstrate this on several occasions before the first OSATS assessment. It is not envisaged that they will successfully complete the assessment at the first attempt

and this should not be seen as failure. The department should nominate an assessor and in some situations discretion may be given to choose their own assessor. Taking consent for the procedure is not part of the assessment; however, the taking of consent must be assessed separately using a mini-CEX. You must retain all OSATS assessment forms, whether satisfactorily completed or otherwise. Review of these forms allows your assessor to see the progress you are making.

There are two parts to the OSATS form. The first is a checklist, which breaks down the procedure into steps, all of which must be successfully completed. The second is a generic technical skills assessment. The generic technical skills, not all of which will be relevant to every OSATS, will form an important part of the assessment process. It is anticipated that, to pass the OSATS, the majority of competences ringed in the middle or to the right of the generic skills assessment list. However, to be signed off for independent practice, the generic skill 'fully understands areas of weakness' within the generic skill of insight/attitude must be consistently ringed.

Trainees will proceed at different rates and the competency levels are the minimum that must be achieved prior to moving to the next stage of training. The OSATS form may be used to assess technical skills at differing levels of complexity; for example, the caesarean section OSATS may be used for assessing competency for a simple caesarean section or a complex caesarean section. The level of complexity should be indicated on the assessment form.

OSATS (OPERATIVE LAPAROSCOPY)

Trainee's name:	

RTN:	**Training Year** (*circle*)**:**	1 2 3 4 5

Date of Assessment:	**Location of Assessment:**

Assessor Name:

Clinical details of complexity/difficulty of case:

Low ☐   Average ☐   High ☐   New Case ☐   Review ☐

	Performed independently	Needs help	Not applicable
	PLEASE TICK THE RELEVANT BOX		
Preparation of the patient			
Ensures correct positioning of the patient, catheterisation and insertion of uterine manipulator			
Patient habitus			
Laparoscopic entry			
Safe use of Veress needle (if used)			
Safe insertion primary port			
Appropriate position of and safe insertion of secondary ports			
Operative procedure			
Maintains good view of operative field			
Uses appropriate instruments for the task			
Knowledge and safe use of energy modalities in laparoscopic surgery			
Identifies important anatomical structures (ureter, internal iliac artery/vein)			
Shows efficiency of movement and demonstrates good three-dimensional spatial awareness			
Appropriate use of assistants (if applicable)			

Examples of minimum levels of complexity for each stage of training

Basic Training   laparoscopic clip sterilisation
Intermediate Training   bipolar diathermy to endometriosis
                               aspiration of fluid form pouch of Douglas
                               aspiration of ovarian cyst
                               Ectopic pregnancy
Advanced   salpingectomy
                               oophorectomy

GENERIC TECHNICAL SKILLS ASSESSMENT

Please ring the candidate's performance for each of the following factors:

Respect for tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments.	Careful handling of tissue but occasionally causes inadvertent damage	Consistently handled tissues appropriately with minimal damage.
Time, motion and flow of operation and forward planning	Many unnecessary moves. Frequently stopped operating or needed to discuss next move.	Makes reasonable progress but some unnecessary moves Sound knowledge of operation but slightly disjointed at times	Economy of movement and maximum efficiency. Obviously planned course of operation with effortless flow from one move to the next.
Knowledge and handling of instruments	Lack of knowledge of instruments.	Competent use of instruments but occasionally awkward or tentative	Obvious familiarity with instruments.
Suturing & knotting skills	Placed sutures inaccurately or tied knots insecurely, and lacked attention to safety.	Knotting and suturing usually reliable but sometimes awkward	Consistently placed sutures accurately with appropriate and secure knots, and with proper attention to safety.
Technical use of assistants Relations with patient and the surgical team	Consistently placed assistants poorly or failed to use assistants Communicated poorly or frequently showed lack of awareness of the needs of the patient and/or the professional team	Appropriate use of assistant most of the time Reasonable communication and awareness of the needs of the patient and/or of the professional team	Strategically used assistants to the best advantage at all times. Consistently communicated and acted with awareness of the needs of the patient and/or of the professional team
Insight/Attitude	Poor understanding of areas of weakness	Some understanding of areas of weakness	Fully understands areas of weakness
Documentation of Procedures	Limited documentation Poorly written	Adequate documentation, but with some omissions, or areas that need elaborating	Comprehensive legible documentation, indicating findings, procedure and

			postoperative management

Based on the checklist and the Generic Technical Skills Assessment, Dr

……………………………………………………

☐ is competent in all areas included in this OSATS

☐ is working towards competence

Needs further help with: * * *Date* *Signed (trainer)* *Signed (trainee)*	Competent to perform the entire procedure without the need for supervision Date Signed Signed

* Delete where applicable, and date and sign the relevant box

Appendix 1   RCPI DOPS - Colonoscopy

FORM O79   GUIDELINES FOR DOPS ASSESSMENTS GASTROENTEROLOGY; COLONOSCOPY, THERAPEUTIC ENDOSCOPY & UPPER GI ENDOSCOPY

Direct observations of practical skills assess the capabilities of an SpR while he/she performs a procedure.   The DOPS is a structured assessment of actual performance.   The basic competencies required for most procedures can be assessed using the list of CORE SKILLS below. SPECIFIC SKILLS relevant to a particular procedure may be added.

Notes for the Assessor:

(Please read)

1. Please ensure the patient (client) is aware that a DOPS is taking place, and that the assessor has been properly introduced.

2. The assessor must be familiar with performing the procedure under assessment, also competent, in case the trainee requires assistance to complete.

3. The assessment should be made under appropriate conditions (e.g. with all equipment, personnel, necessary to support) and in a suitable environment.

4. The task to be undertaken should be of a degree of difficulty appropriate to the operator's capabilities.

5. Marks are given for each category for well below expectations, below expectations, borderline, meets expectation, above expectation and well above expectation for stage of training. Borderline and under are considered unsatisfactory and if any of these marks are given the assessor **must** explain in the space provided for comments. **The marks given should fairly represent the trainee's performance in this evaluation as expected for their stage of training.**

6. The assessor **must** also give feedback to the trainee on the performance assessment just completed, dealing fully with any deficiency identified. Both the **trainer** and the trainee sign the form to confirm that this has taken place, and space is provided here for any additional comments from the trainee and the trainer.

7. The fully completed form is handed to the trainee. It is their responsibility to keep a copy for their portfolio.

Competencies to be assessed:

Understanding of Procedure: Relevant anatomy; purpose, indications, contra-indications; outcomes, risks, complications; choice of methods available.

Consideration for the Patient: Gives reassurance, minimises discomfort, explains procedure fully; confirms informed consent obtained.

Preparation: First re-checks all relevant details correct. Safety check; instrumentation, equipment (drugs); positioning; cleansing/aseptic technique; sedation, analgesia, anaesthesia confirmed.

Professional/technical ability: Dexterity, accuracy, efficiency; obtains, interprets diagnostic material/information; informs, directs staff courteously; recognises own limitations; manages risk.

Post-Procedure: Completes documentation; regulates recovery phase, observations; anticipates/deals with complications. Informs/counsels patient/relatives.

Overall ability to perform Procedure: Ability to complete/undertake procedure; technical abilities as demonstrated; appropriately confident, team/ leadership skills.

FORM O79 DOPS Assessment Colonoscopy							
Trainee's Name:							
RTN:		**Training Year:**	**1**	**2**	**3**	**4**	**5**
Date of Assessment:		**Location of Assessment:**					
Procedure Undertaken:		Colonoscopy					
Complexity:	Low		Average			High	

1. Well Below Expectation	**2.** Below Expectation	**3.** Borderline
4. Meets Expectation	**5.** Above Expectation	**6.** Well Above Expectation

1. Obtained informed consent: ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   Consent should include procedural information, risk and complications explained, co-morbidity, sedation and outcomes.
2. Preparation - Safety and Sedation: ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6
3. Endoscopic Skills   3.1 Checks endoscope function before intubation   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.2 Performs PR   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.3 Maintains luminal view/inserts in luminal direction   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.4 Demonstrates awareness of patient's consciousness   and pain during the procedure and takes appropriate action   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.5 Uses torque steering   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.6 Uses distension, suction & lens washing appropriately   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.7 Uses position change and abdominal pressure to aid   luminal views   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6   3.8 Completes procedure in reasonable time.   ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4. Diagnostic & Therapeutic Ability

4.1 Adequate mucosal visualisation
  ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.2 Recognises caecal landmarks or incomplete examination

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.3 Accurate identification & management of pathology

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.4 Use diathermy and therapeutic techniques appropriately

and safely
  ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.5 Recognises & manages complications appropriately

  ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

5. Professionalism and consideration for the patient during the procedure

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

6. Overall competence in performing procedure

  ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Comments:

Signed by Trainer:	Date:
Signed by Trainee:	Date:

Appendix 2   RCPI DOPS – Therapeutic Endoscopy

FORM 080 DOPS Assessment Therapeutic Endoscopy						
Trainee's Name:						
RTN:		**Training Year:**	**1**	**2**	**3**	**4**	**5**
Date of Assessment:		**Location of Assessment:**				
Procedure Undertaken: Therapeutic Endoscopy						
Complexity:	Low		Average		High	

1. Well Below Expectation	**2.** Below Expectation	**3.** Borderline
4. Meets Expectation	**5.** Above Expectation	**6.** Well Above Expectation

---

1. Obtained informed consent:

○ **N/A**   ○ **1**   ○ **2** ○ **3** ○ **4** ○ **5** ○ **6**

Consent should include procedural information, risk and complications explained, co-morbidity, sedation and outcomes.

2. Preparation - Safety and Sedation:

○ **N/A**   ○ **1**   ○ **2** ○ **3** ○ **4** ○**5** ○ **6**

3. Endoscopic Skills

3.1 Skilful endoscopic handling to enable diagnosis and

treatment

○ **N/A**   ○ **1**   ○ **2** ○ **3** ○ **4** ○**5** ○ **6**

3.2 Demonstrates awareness of patient's consciousness and

comfort during the procedure and takes appropriate actions

○ **N/A**   ○ **1**   ○ **2** ○ **3** ○ **4** ○ **5** ○ **6**

3.3 Completes procedure in reasonable time

○ **N/A**   ○ **1**   ○ **2** ○ **3** ○ **4** ○ **5** ○ **6**

4. Diagnostic & Therapeutic Ability

4.1 Adequate identification & visualisation of therapeutic target

○ **N/A**   ○ **1** ○ **2** ○ **3** ○ **4** ○**5** ○ **6**

4.2 Chooses an appropriate therapy from the available

range of options

○ **N/A**   ○ **1** ○ **2** ○ **3** ○ **4** ○ **5** ○ **6**

4.3 Safe and effective management of pathology within

limits of pathology, using therapeutic techniques appropriately

○ **N/A**   ○ **1** ○ **2** ○ **3** ○ **4** ○ **5** ○ **6**

4.4 Recognises & manages complications appropriately

○ **N/A**   ○ **1** ○ **2** ○ **3** ○ **4** ○ **5** ○ **6**

5. Professionalism and consideration for the patient during the procedure ○ **5** ○ **6**	○ **N/A** ○ **1** ○ **2** ○ **3** ○ **4**
6. Overall competency in performing procedure ○ **5** ○ **6**	○ **N/A** ○ **1** ○ **2** ○ **3** ○ **4**

Comments:

Signed by Trainee:		**Date:**
Signed by Assessor:		**Date:**

Appendix 3   RCPI DOPS – Upper GI Endoscopy

FORM 080 DOPS Assessment Therapeutic Endoscopy						
Trainee's Name:						
RTN:	**Training Year:**	**1**	**2**	**3**	**4**	**5**
Date of Assessment:	**Location of Assessment:**					
Procedure Undertaken: **Therapeutic Endoscopy**						
Complexity:	Low		Average		High	

1. Well Below Expectation	**2.** Below Expectation	**3.** Borderline
4. Meets Expectation	**5.** Above Expectation	**6.** Well Above Expectation

---

1. Obtained informed consent:

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Consent should include procedural information, risk and complications explained, co-morbidity, sedation and outcomes.

2. Preparation - Safety and Sedation:

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3. Endoscopic Skills

3.1 Checks endoscope function before intubation

 ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3.2 Intubates the oesophagus under direct vision

 ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3.3 Maintains luminal view

 ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3.4 Demonstrates awareness of patinet's consciousness and

comfort during the procedure and takes appropriate actions

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3.5 Used distension, suction & lens washing appropriately

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3.6 Passes the scope into the second part of the duodenum

 ○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

3.7 Completes procedure in reasonable time

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4. Diagnostic & Therapeutic Ability

4.1 Adequate mucosal visualisation

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.2 Recognises & Notes the position of the gastro-oesophageal

junction and is appropriately orientated with the stomach
and duodenum

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.3 Accurate identification & management of pathology

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4.4 Uses diathermy and therapeutic techniques appropriately
and safely

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

4. 5 Recognises & manages complications appropriately

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

**5. Professionalism and consideration for the patient
during the procedure**

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 .○ 5 ○ 6

---

6. Overall competency in performing procedure

○ N/A   ○ 1 ○ 2 ○ 3 ○ 4 .○ 5 ○ 6

Comments:

Signed by Trainee:	**Date:**
Signed by Trainer:	**Date:**
Signed by Assessor:	**Date:**

Appendix 4   REC approval letter

ROYAL COLLEGE OF
PHYSICIANS OF IRELAND

Frederick House
19 South Frederick Street, Dublin 2
Telephone: +353 1 863 9700
Facsimile: +353 1 672 4707
Website: www.rcpi.ie

Aileen Barrett
Education Department,
RCPI
Frederick House,
19 South Frederick Street
Dublin 22

July 8 2013

Dear Ms Barrett

Re: An audit of compliance with mandatory workplace- based assessment requirements in postgraduate medical training.

Dear Ms Barrett,

Thank you for submitting your audit to the research ethics committee for review.

The Research Ethics Committee's opinion is favourable.

It is however recommended that all trainers and trainees be advised that the audit is being conducted.

The committee would like to wish you every success in undertaking this audit.

Yours Sincerely

Davide De La Harpe 04021

_____
Dr Davida De La Harpe

CHAPTER 2   FIGURES

Fig 1   Data extraction tool

DATE OF AUDIT: _____   AUDIT PERIOD: July 2012- July 2013   Trainee code: _____

SPECIALTY _____   Year of Training   1   2

MINIMUM REQUIREMENTS:

DOPS/OSATS _____   Mini-CEX _____   CbD   _____

(+/-Dual requirements)

WBA	PROCEDURE/TOPIC	NO. WEEKS IN POST	EVIDENCE OF SPECIFIC FEEDBACK	SPECIFIC GOALS	ANY DOMAIN BELOW EXPECTATION/BORDERLINE	FOLLOW-UP WBA FOR COMPETENCIES BELOW EXPECTATION/BORDERLINE
DOPS/OSATS 1						
DOPS/OSATS 2						
DOPS/OSATS 3						
DOPS/OSATS 4						
CbD 1						
CbD 2						
CbD 3						
CbD 4						
Mini-CEX 1						
Mini-CEX 2						
Mini-CEX 3						
Mini-CEX 4						

CHAPTER 3   APPENDICES

Appendix 1   REC approval letter

ROYAL COLLEGE OF PHYSICIANS OF IRELAND

Frederick House
19 South Frederick Street, Dublin 2
Telephone: +353 1 863 9700
Facsimile: +353 1 672 4707
Website: www.rcpi.ie

Ms Aileen Barrett
Education Specialist
Education and Professional Development
Royal College of Physicians of Ireland
Frederick House
19 South Frederick Street
Dublin 2.

07 October 2013

**Re: Trainer and trainee perceptions and experiences of workplace-based
assessments as a learning tool**

Dear Ms. Barrett,

Thank you for submitting your research proposal to the Research Ethics Committee at the
Royal College of Physicians and for attending the meeting on the 19th of September.

The Research Ethics Committee's opinion is **favourable**.

Please note that the committee requires the following to be submitted.

1. A yearly update report.
2. A premature termination report (If research is stopped early).
3. A completion report.

Also in the event of any adverse event occurring in the course of this research (e.g. breach
of confidentiality), the committee should be informed as soon as practicable via the chair of
the committee.

The committee would like to wish you every success with this project.

Yours Sincerely,

PP Claire Davenport

_____

Dr. Una Fallon
Acting Chair RCPI REC

Appendix 2   Interview guide - trainers

- Tell me about your experiences of WBA as a trainer
- What role, if any, do you think WBAs play in trainee learning? Examples
- In terms of feedback, do WBAs have a role to play in delivery?
- What are the barriers to use as a learning tool in clinical practice?
    - Prompt: are there any other challenges?
- What are the facilitators to its use as a learning too?
- What are your thoughts on the three tools currently in use – the Mini-CEX, DOPS, CbD – as learning tools?
- Do you think the impact of WBAs on learning could be improved? How?

Appendix 3   Interview guide – trainees

- Tell me about your experiences of WBA as a trainee
- What role, if any, do you think WBAs play in learning as a trainee? Can you give me examples of where it did/did not help in learning?
- In terms of feedback, do WBAs have a role to play in delivery? What is your experience of feedback following a WBA as a trainee? Examples
- What are the barriers to use as a learning tool in clinical practice?
    - Prompt: are there any other challenges?
- What are the facilitators to its use as a learning too?
- What are your thoughts on the three tools currently in use – the Mini-CEX, DOPS, CbD – as learning tools?
- Do you think the impact of WBAs on learning could be improved? How?

Appendix 4   Sample transcript

Trainee 3   AB= Aileen Barrett (interviewer)   Te = trainee

AB: So really my first question is around your experiences of workplace based
         assessment as a trainee, what ones you had and really were they positive negative
         or mutual experiences?

Te: Ok so the ones, well certainly on the paper based system we find that as with
         anything form filling comes to the very end of the year so you kind of try, honestly
         speaking there isn't an opportunity or we don't do it on a continuous basis, it's
         really done at the tail end of the year where we're coming to assessment so there
         is no continual assessment as such it's only done towards the end and then we're
         just throwing together cases that we can recollect and the signing off is more of a
         formality rather than an actual endeavour

AB: So have you had experiences with all 3 in GIM or just the mini CEX and the case
         based discussions?

Te: So there's case based discussions, mini CEX and what's the other?
AB: DOPS

Te: DOPS, I've done all of them yeah.

AB: So you've done all of them ok, so as a learning tool, the tools in themselves, did
         they facilitate any kind of learning for you?

Te: I think for the direct observed procedures, many of those ones that you require for
         different areas, I would be come into my mind, competent in them at SHO level so
         some of the more specially related procedures then you do get signed off at a
         registrar level, now I was a registrar prior to the specialist training programme so
         there was no formal assessment.

AB: So it tends to frontload though towards the earlier part of training?

Te: Yeah you know so for all the procedures that we do in my speciality, sort of
         insertion of chest drain, hands on specialist procedures you become proficient in
         those in the first year of training and subsequently you carry on unobserved for the
         rest of your training because your providing a service.

AB: Once you're deemed to be competent?

Te: Yes, yeah so, now having said that I was fully observed for the first number of
         procedures and for example, the bronchoscopy I had a hundred or so, in one

hospital there was a hundred+ bronchoscopies supervised by a consultant which you don't get in every institution so yeah you became proficient at it but there was no formalised recognition of that training.

AB: Nowhere to put it.

Te: Nowhere to put it, yeah.

AB: With the 3 of them and the number of times you've done them over the last couple of years does anyone stick out as a positive learning experience, is there anyone that stands out oh actually because I did that I learned something.

Te: Am probably not no, I'm just thinking back, like certainly if you're thinking of the mini CEX really didn't, I don't think that was formally done, I think that's more of a, if you think of the, or the case base discussions again you're going through a clinical scenario in detail with your consultant or what not and that tends to happen at clinic quite regularly anyway, it your fortunate to have somebody who's interested in training in that regard so I don't know if a form helps that because you don't bring forms to clinic, you don't carry them around with you. You're not going to, your consultant isn't going to sit down with a computer with you and go through it with you with you either.

AB: Yeah, did you get feedback through your training without them?

Te: I received feedback without them yes, so without the forms I would have had feedback, not frequent intervals but you know you might you know 6 months or 3 months depending on your trainer, would sit down with you, go for a cup of coffee and say you're getting on fine or sometimes not so much you know.

AB: Is there anything, are there any changes to either the tools or the practices by which we use them and you've mentioned that a lot of the time they're left until the end of the year so the learning is probably limited, are there any changes to either the tools or the practices that would make them more effective as a learning experience for the trainees?

Te: I think from a learning experience the, there's a difficulty trying to segregate the learning experience from the clinical workload that you have to get through in a particular day, consultants are quite busy and different things also the teams are quite busy and you're under pressure to look after patients perhaps who are not, in terms of speciality training, they're maybe not in your speciality or what not so perhaps that dilutes the focus in terms of your day to day training in that area obviously. But if you think of, let's say, if you were to take a particular procedural skill which is important perhaps more formalised training at the earlier stages of your career. So I mentioned bronchoscopy earlier on, I know that there are some bronchoscopy courses that maybe tailored to other things, you know in terms of

cardiology, gastroenterology, neurology, that maybe particular skills that you should acquire at a very early stage of your training that are formalised over, that they're taught formally and perhaps in a standardised fashion, I'm sure there's standardised operating procedures for all of these anyway but that they're thought early, that you attend this and then perhaps you're assessed at a later stage, yeah whether you get assessed externally rather than internally you know.

AB: So what is your understanding of the purpose of the workplace based assessment, why do you think they're there?

Te: My understanding of it is so that the training that we receive is accredited internationally so there has to be some way of saying ok they have done this or accomplished this whether it is actually formative for your career and development I'm not sure if it is, I think its just a process that you do to get signed off.

AB: Ok so that leads me to the question then as a learning tool should these workplace based assessments remain as a mandatory feature of training?

Te: I think that they're artificial and perhaps they aren't constructive to the end result but I think the goal should be maintained so as I mentioned you could have a case based discussion or something like that could be formalised into perhaps a, you know a specialist registrar get together where you do case studies or particularly, I think that would be more sort of peer-centred case-based discussions or what not rather than, you know that are chaired by a consultant or a group of consultants in that area rather than just a sheet of paper and then that's a better learning experience to go through so that might be one.

AB: So the concept should remain.

Te: The concept should remain so you'd have structured perhaps didactic teaching in a procedural skill structured case discussion in a small group with a group of your peers like at a study day or something like that. That you know you'd each prepare a case that you'd seen in the past 6 months or 3 months or whenever and then sit around and discuss it and that way at least you're contributing to other people's learning and not just your own.

AB: Yeah exactly, so possibly changes to the process by which we use them, so I'm thinking as well about the timing and what you said earlier about the fact that they're left until the end of the year, and there isn't much time to obviously effect any change to your performance, is there a way, should we be introducing them as in…..

Te: Continuous assessment.

AB: Exactly, should we say right now we have one mini CEX per year or training programme, it's per year for most of them, if we said it was one per quarter would that stage it better?

Te: Again witnessing the surgical, you know having colleagues who did surgical training, they have continual month on month sort of computer based assessments where they have to complete a particular element of their training on line or what not, I don't know how receptive people would be having to fill in these things on deadlines, again it comes across as a form filling exercise without, you know you go through it, you put in the information but you're not actually processing it.

AB: But would you learn from it?

Te: I don't think, as a learning exercise it's a good learning exercise and probably…

AB: That would increase the impact?

Te: Yeah, yeah and probably putting people through it on a continual basis may actually make more negative associations with it…..

AB: For a while.

Te: For a while yeah, so perhaps, you know I think meeting up and discussing it on a more, at a study day or something like that, going through a number of cases that you would attend and present a couple of cases, at least when you're presenting to other people you'll process the information that's important and learn from it and then be able to impart it to other people.

AB: And would you be looking in that situation for sort of peer feedback then, and say have the case based discussion format or form and the rating kind of process and sort of do it in a peer group?

Te: Yeah I think that would be better because you'd be, you'd have to think a bit more about it and process it a bit more, you're not processing it when you're putting it onto a sheet of paper to be honest, your just and your thinking back, oh that was an interesting case, we did discuss it in a clinic you know these are the points that we raised, this is what I remember but then sure I can't remember any case based thing that I did.

AB: None of the feedback stands out?

Te: None of the feedback stands out.

AB: Ok I think that's actually it, I think you've actually talked through a lot of the questions I had already, so if I was to just summarise what you were saying then so

you think that as a concept and the concept of formative assessment being to provide feedback is a good one, that should stay, the process needs to tweaked so that it actually has a better impact on learning, be that staging it through the year or the mechanisms by which it's used and including maybe peer feedback in that.

Te: Yeah perhaps again some of the onus should be shifted to the trainers as well rather than the trainees, you know if, you know at the moment sometimes you're chasing them down, sometimes you don't get any feedback whatsoever so if a trainer wants to be a trainer they need to kind of step up to it and perhaps....

AB: Have you ever had a trainer refuse to do one?

Te: No, no, I've never had one refuse to do one no.

AB: No. It's just hard to get them tied down.

Te: Or not to volunteer either to be honest so you are chasing them a bit towards the end of the year to do it and they're happy to do it, it's just perhaps if you're talking about say, everybody is meant to sit down every 3 months and have their training reviewed, that doesn't really happen in the real world perhaps through no fault of either group but if it was to be formalised perhaps the onus should be switch to the trainer, as it is in other jurisdictions that the trainers are there to, the burden of teaching lays with the trainers, which I think is reasonable.

AB: Fair point, I think that's it, did you have anything else to add?

Te: No, no, I'm happy with that.

AB: And I suppose the purpose of this is to start making changes but just to make sure that they are both trainer- and trainee-informed.

Appendix 5   Final trainer template

1. Experiences
    a. Trainee engagement/responsibility/initiative
        i. Poor trainee engagement
        ii. Lots of opportunities
    b. Drivers for WBA
        i. Institution-driven
            1. Imposition/trainee-centred initiative
    c. Trainer role and integration of WBA in own context
        i. Workload
        ii. NSD – additional engagement in teaching and knowing 'evidence' for WBA
    d. How WBAs are used
        i. Purpose of WBA
            1. Purpose of WBA in general
                a. Assessment of performance vs assessment for learning
        ii. Trainee responsibility
            1. Trainers engaged when asked
        iii. Timing
            1. End-of-year
            2. WBA as a 'tick-box'
            3. Retrospectively
        iv. Tools
            1. Purpose of individual tools
            2. Familiarity with tools
            3. Design
        v. Role of WBA in feedback delivery
            1. Structure
            2. Negative and positive feedback can be facilitated
    e. Concurrent experience of new technology/innovation (ePortfolio)
        i. Issues with efficiency
        ii. Delayed sign-off – impact on feedback
    f. Issues with authenticity
        i. Linked to understanding of purpose of WBA as assessment of performance

2. Perceptions of learning value
    a. Current value
        i. Feedback

1. Delivery of positive and negative
2. Role of trainer in day-to-day practice negates need for WBA
   ii. Facilitate reflection on practice
b. Potential (not-yet-realised) value
   i. 360 MSF as potential source of feedback
      1. Issues with responsibility

1. Experiences
    a. Purpose of WBA
        i. Fulfilling requirements/ticking boxes
            1. Associated with end-of-year assessment/logbook completion
        ii. Assessment of competence/sign-off
            1. Academy-driven
        iii. Familiarity with tools
            1. Mini-CEX and CbD perceived as part of routine practice
            2. Mini-CEX related to case presentation (hx/exam)
    b. Use of WBA
        i. Retrospective recall of case++
        ii. Use of the term 'formal'
        iii. Trainees 'teaching' trainers to use WBA
    c. Authenticity
        i. Competing service and learning demands
        ii. Relevance
            1. Variation in perceptions of relevance of specific tools within and between specialties
    d. Responsibility
        i. Trainer vs trainee responsibility
        ii. Relationships and impact on responsibility/initiation of WBA
        iii. Training body responsibility to teach trainers and trainees

2. Perceptions of learning value
    a. Realised
        i. Forces trainer observation
    b. Potential
        i. Feedback
            1. Diagnostic assessment
            2. Affirming good practice vs 'deficit focus'
            3. 'Anchoring'
                a. Situating performance and creating goals
            4. Feedback from other professionals

Figure 1-4   **Template development photos** (trainee interviews)

Fig 1: Preliminary coding scripts batch 1

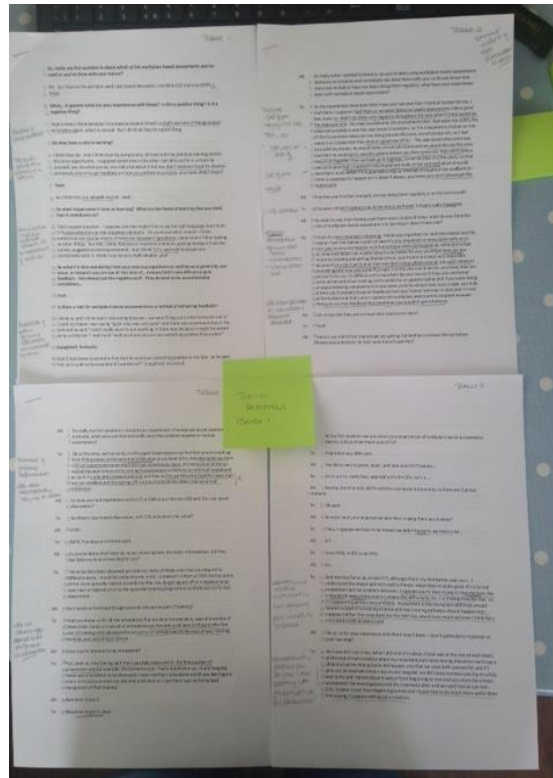**Fig 2: Coding interview 1 & 2**

Fig 3: 'Clustering' batch 1

Fig 4: 'Clustering' all interviews

Appendix 1   BEME protocol approval email

2014 11 27_ SARA HEJRI to AILEEN BARRETT regarding the revised protocol

Sara Mortaz Hejri <sa_mortazhejri@razi.tums.ac.ir>

Reply all|

To:
Aileen Barrett;

Cc: beme@dundee.ac.uk
Thu 27/11/2014 14:47
PhD

You replied on 28/11/2014 11:00.

Action Items

Dear Aileen,

We were delighted to receive and review your revised protocol.

Since the referees' concerns and comments have been fully addressed, I am glad to inform you that the Tehran BICC decided to consider the protocol as acceptable in its present form.

Your protocol has been now uploaded to the BEME website and we would be grateful if you could confirm that the information is correct.

http://www.bemecollaboration.org/Reviews+In+Progress/work-based+assessments/

You can progress to the next stage. Please be aware that regular updates are required from you on the progress of the review.

On behalf of the Tehran BICC, I would like to thank you and your colleagues for your cooperation during the review process. I wish you success in conducting your review, and hope your project goes well.

If you come up with any question please feel free to contact us.

Best regards,

Sara

--
Sara Mortaz Hejri,
MD, MSc, PhD candidate,
Medical Education Department, School of Medicine,
Tehran University of Medical Sciences,
Deputy Director of Educational Development Office, School of Medicine
Head of TUMS-BICC (BEME International Collaborating Center)
Tehran,
Iran

Appendix 2   Sample MEDLINE search

Search strategy:
1. ((workplace-based or work-based or workplace based or work based) adj (feedback or assessment* or evaluation*)).tw.
2. ((multisource or multi-source or 360-degree or 360 degree) adj (feedback or assessment* or evaluation*)).tw.
3. (formative adj (feedback or assessment* or evaluation*)).tw.
4. (assessment adj2 learning).tw.
5. (Clinical Evaluation Exercise or Mini-CEX or mCEX).tw.
6. (Mini Peer Assessment Tool or Mini-PAT).tw.
7. case based discussion.tw.
8. (Direct Observation of Procedural Skills or procedure based assessment).tw.
9. Structured Assessment of Technical Skills.tw.
10. or/1-9
11. Clinical Competence/
12. (feedback or assessment* or evaluation*).tw.
13. 11 and 12
14. Education, Medical, Graduate/
15. (postgraduate* or post-graduate* or resident* or junior*).tw.
16. 14 or 15
17. or/1-9,13
18. 10 and 16
19. 16 and 17
20. exp Physicians/
21. (physician* or doctor* or medical).tw.
22. 20 or 21
23. 18 and 22
24. 19 and 22
25. limit 24 to yr="1995 -Current"
26. limit 25 to (dutch or english or french or german)
27. (comment or editorial or letter).pt.
28. 26 not 27
29. exp Animals/ not (exp Animals/ and Humans/)
30. 28 not 29

CHAPTER 5   APPENDICES

Appendix 1   PRISMA checklist

TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1; title as per BEME guidelines
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 3 (abstract guidelines for *Medical Teacher*)
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 4 – 7 (page 4 summarises the 'problem-gap-and-hook')
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Research questions: Page 6; PICOS: Page 9
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Page 8: publication referenced
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 9
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 8

Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Medline search: Appendix 3); search log of other databases: Appendix 4 (supplementary material)
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 9-10
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 12
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 9: PICOS. No funding sought
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 11: tools used to evaluate methodological quality (MQ). MQ evaluation in Appendix 8 and 9
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Page 26

Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Page 15 -16: evaluation of MQ specifically addressing risk of bias assessment
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A

RESULTS				
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Page 13-14
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Page 17-22
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Page 15-16 (see No. 15 above)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A – narrative synthesis conducted
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Page 32-34
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 36
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Page 32-34
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Page 37

Appendix 2   STARLITE reporting checklist

Sampling strategy	Strategy development and peer review process documented
Type of study	BEME review
Approaches	detailed in search strategy - electronic sources, grey literature, database alerts, hand searching of key journals
Range of years	1995-2015
Limits	English, French, German and Dutch; humans only
Inclusions and exclusions	PICO details provided
Terms used	Search terms provided in appendix
Electronic sources	CINAHL, MEDLINE, BEI, AEI, EMBASE, PsycInfo, Cochrane, Science Direct

Appendix 3   MEDLINE search (following peer review)

<u>Database:</u>

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present.

<u>Search strategy:</u>

1. ((workplace-based or work-based or workplace based or work based) adj (feedback or assessment* or evaluation*)).tw.

2. ((multisource or multi-source or 360-degree or 360 degree) adj (feedback or assessment* or evaluation*)).tw.

3. (formative adj (feedback or assessment* or evaluation*)).tw.

4. (assessment adj2 learning).tw.

5. (Clinical Evaluation Exercise or Mini-CEX or mCEX).tw.

6. (Mini Peer Assessment Tool or Mini-PAT).tw.

7. case based discussion.tw.

8. (Direct Observation of Procedural Skills or procedure based assessment).tw.

9. Objective Structured Assessment of Technical Skills.tw.

10. or/1-9

11. Clinical Competence/

12. (feedback or assessment* or evaluation*).tw.

13. 11 and 12

14. Education, Medical, Graduate/

15. (postgraduate* or post-graduate* or resident* or junior*).tw.

16. 14 or 15

17. or/1-9,13

18. 10 and 16

19. 16 and 17

20. exp Physicians/

21. (physician* or doctor* or medical).tw.

22. 20 or 21

23. 18 and 22

24. 19 and 22

25. limit 24 to yr="1995 -Current"

26. limit 25 to (dutch or english or french or german)

27. (comment or editorial or letter).pt.

28. 26 not 27

29. exp Animals/ not (exp Animals/ and Humans/)

30. 28 not 29

Appendix 4   Database search log

Date	February 2015
Research Topic	Workplace-based assessment
Review Title	A BEME systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees
Question	*Primary Review Question* Can workplace-based assessment be used to identify and remediate poor performance among postgraduate medical trainees? *Secondary Question* What features of workplace-based assessment tools and/or factors associated with WBA methods and utilisation primarily contribute to the usefulness of WBA in identifying or remediating poor performance among postgraduate medical trainees?
Search Strategy	**Key concepts** 1. Workplace-based assessment 2. Postgraduate medical education/training 3. Poor performance/ underperformance 4. Remediation	**Synonyms/alternative terminology (consider regional variations here also) – combine using OR** • Workplace-based assessment OR formative assessment OR feedback OR evaluation OR assessment for learning • Postgraduate medical education OR postgraduate medical training • Postgraduate medical trainee* OR postgraduate medical student* OR trainees OR resident* OR junior doctor OR physicians-in-training • Trainee*-in-difficulty OR doctor*-in-difficulty OR trainee*-in-trouble OR doctor*-in-difficulty OR academic difficulty OR performance deficit • Mini-clinical evaluation exercise OR mini-CEX • DOPS OR direct observation of procedural skills • Objective structured assessment of technical skills OR OSATS • Mini-PAT OR mini-peer assessment tool

		• Case-based discussion OR CBD • Multi-source feedback OR 360° feedback OR MSF
Limits **Time / gender / age /** **geography / ethnic & cultural /** **[language]**	• 1995 – current • English - French - Dutch - German	
Inclusion criteria		
	Population	• Postgraduate medical trainees • Postgraduate surgical trainees
	Intervention	The interventions which will be considered for this review are those which involve the use of workplace-based assessment either routinely (e.g. as a component of clinical rotations), or in relation to poor performance (e.g. confirmation of poor performance) We will include studies that describe or evaluate the use of WBA within the context of: • Routine or targeted use of WBA • Trainee-led or trainer-led WBA • Single or multiple-use of WBA tools • Management or remediation of poor performance for knowledge, skills and attitudes
	Outcomes	Studies that describe/report outcomes related to identification/remediation At least one of the following outcomes must be reported as being specifically resultant from engagement in a compone part of a WPA assessment process: *Individual* • Number of trainees identified as poorly performing through the use (either routine or targeted) of a WBA process • Progression/remediation statistics

		• Changes in trainee performance (knowledge, skills, attitudes etc.) • Trainee satisfaction *Practice* • Changes in implementation methods e.g. non-routine to routine • Implementation of new/differing WBA tools *System* • Changes in system-wide implementation of WBA tools or methods e.g. throughout a deanery
	Research Design	Studies which provide primary data for any of the outcomes above including, but not limited to, the following designs: • Experimental and/or observational studies • Randomised and non-randomised studies • Prospective or retrospective cohort studies • Qualitative • Descriptive
Exclusion criteria	Population • Non-medical trainees • Medical students (undergraduate and graduate-entry programmes) • Studies not involving humans • Studies in medicinal areas **not** related to humans (e.g. veterinarian studies) • Studies not involving physicians Research Design • Studies that do not report an outcome including commentaries, letters and editorials • Reports published only in dissertation and abstract format	

Resources to be searched	
Databases	▪ Medline ▪ CINAHL ▪ British Education Index ▪ EMBASE ▪ ERIC ▪ Australian Education Index ▪ BEME published reviews, Cochrane, DARE ▪ PsycINFO ▪ Science Direct
Grey literature	Reference lists of included studies and review articles http://www.ncbi.nlm.nih.gov/pubmed/21833989
	Contact with prominent authors in the field of workplace-based assessment for expert recommendations and guidance and to identify unpublished (including doctoral theses), recently published or ongoing studies relevant to this review. Prominent authors are defined as those having published previous relevant reviews multiple WBA studies
	Conference presentations from the Association of Medical Education in Europe, Association for the Study of Medical Education, International Conference on Residency Education and Canadian Conference on Medical Education will be searched for relevant abstracts from 1995 forward or the inaugural year of the conference as applicable.
	We will also conduct a citation search on Web of Science looking for studies citing any of the included articles

Search Log overview

Resource name	Details of search	Number of hits	Details of key resources and actions taken
MEDLINE	Platform: OVID Medline® in-process & other non-indexed citations and OVID Medline ® 1946-present Search History No.1	2277	Saved all to EndNote as Medline#1
CINAHL	Platform: EBSCO Host CINAHL Plus with Full text Search history No.2	346	Saved all to EndNote as CINAHL#1
ERIC	Platform: EBSCO Host Search History No.3	40	Saved all to EndNote as ERIC#1
PsycInfo	Platform: EBSCO Host Search History No. 4	3591	Saved all to EndNote as PsycINFO#1
British Education Index	Platform: ESBCO Host Search History No.5	92	Saved all to EndNote as BEI#1
EMBASE	Platform: Elsevier Search history No. 6	110	Saved all to EndNote as EMBASE#1
Science Direct	Platform: Elsevier Search history No. 7	359	Saved all to EndNote as SciDi#1
Australia Education Index	Platform: ProQuest Search History No. 8	221	Saved all to EndNote as AEI#3
Theses	ProQuest Dissertations Theses A&I	5	Saved all to EndNote as ProQuest Theses A&I
	ProQuest Dissertations Theses UK	0	
	DART Europe e-theses portal	0	
BEME, Cochrane	Cochrane database	25	Saved all to Endnote as Cochrane#1
	BEME	0	

Search Histories

1. MEDLINE #1 25th March 2015

Search History (30 searches) *(close)*

# ▲	Searches	Results	Search Type	Actions
1	((workplace-based or work-based or workplace based or work based) adj (feedback or assessment* or evaluation*)).tw.	192	Advanced	Display More ≫
2	((multisource or multi-source or 360-degree or 360 degree) adj (feedback or assessment* or evaluation*)).tw.	277	Advanced	Display More ≫
3	(formative adj (feedback or assessment* or evaluation*)).tw.	1139	Advanced	Display More ≫
4	(assessment adj2 learning).tw.	869	Advanced	Display More ≫
5	(Clinical Evaluation Exercise or Mini-CEX or mCEX).tw.	144	Advanced	Display More ≫
6	(Mini Peer Assessment Tool or Mini-PAT).tw.	12	Advanced	Display More ≫

7	case based discussion.tw.	84	Advanced	Display More ≫
8	(Direct Observation of Procedural Skills or procedure based assessment or DOPS).tw.	456	Advanced	Display More ≫
9	Objective Structured Assessment of Technical Skills.tw.	123	Advanced	Display More ≫
10	or/1-9	3101	Advanced	Display More ≫
11	Clinical Competence/	67409	Advanced	Display More ≫
12	(feedback or assessment* or evaluation*).tw.	1520968	Advanced	Display More ≫
13	11 and 12	15585	Advanced	Display More ≫
14	Education, Medical, Graduate/	22299	Advanced	Display More ≫

15	(postgraduate* or post-graduate* or resident* or trainee* or junior*).tw.	152728	Advanced	Display More ≫
16	14 or 15	167940	Advanced	Display More ≫
17	or/1-9,13	17877	Advanced	Display More ≫
18	10 and 16	704	Advanced	Display More ≫
19	16 and 17	4770	Advanced	Display More ≫
20	exp Physicians/	88978	Advanced	Display More ≫
21	(physician* or doctor* or medical).tw.	1063146	Advanced	Display More ≫
22	20 or 21	1097671	Advanced	Display More ≫

23	18 and 22	384	Advanced	Display More ≫
24	19 and 22	2599	Advanced	Display More ≫
25	limit 24 to yr="1995 -Current"	2379	Advanced	Display More ≫
26	limit 25 to (dutch or english or french or german)	2303	Advanced	Display More ≫
27	limit 26 to (comment or editorial or letter)	25	Advanced	Display More ≫
28	26 not 27	2278	Advanced	Display More ≫
29	exp Animals/ not (exp Animals/ and Humans/)	4002853	Advanced	Display More ≫
30	**28 not 29**	**2277**	**Advanced**	Display More ≫

2. CINAHL # 1 July 15th 2015

Search ID#	Search Terms	Search Options	Actions
S14	S12 AND S13	**Search modes** - Boolean/Phrase	**View Results** (346) **View Details** **Edit**
S13	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9	**Search modes** - Boolean/Phrase	**View Results** (5,455) **View Details** **Edit**
S12	S10 AND S11	**Search modes** - Boolean/Phrase	**View Results** (10,148) **View Details** **Edit**
S11	medic* OR surg* OR medical education OR medical training	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (319,829) **View Details** **Edit**
S10	postgraduate* OR post-graduate* OR resident* OR junior* OR trainee*	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (28,700) **View Details** **Edit**
S9	OSATS OR objective structured assessment of	**Limiters** - Published Date: 19950101-20150731;	**View Results** (22)

		technical skills	Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Details** **Edit**
S8	case based discussion OR case-based discussion	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (28) **View Details** **Edit**
S7	TX mini-PAT OR TX mini peer assessment tool	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (13) **View Details** **Edit**
S6	TX direct observ* of procedural skills OR TX DOPS OR TX technical skills	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (2,052) **View Details** **Edit**
S5	TX mini-cex OR mini clinical evaluation exercise	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review	**View Results** (89) **View Details** **Edit**

		Search modes - Boolean/Phrase	
S4	TX assessment N2 learning	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (1,913) **View Details** **Edit**
S3	TX formative feedback OR TX formative assessment* OR TX formative evaluation	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (1,780) **View Details** **Edit**
S2	(multisource OR multi-source OR 360-degree OR 360 degree) N2 (feedback OR assessment* OR evaluation*)	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (118) **View Details** **Edit**
S1	(workplace-based N2 (assessment* OR feedback OR evaluation*)) OR (workplace based N2 (assessment* or feedback OR evaluation*)) OR (work-based N2 (assessment* or feedback OR evaluation*)) OR (work based N2 (assessment* OR feedback OR evaluation*))	**Limiters** - Published Date: 19950101-20150731; Human; Language: Dutch, English, French, German; Publication Type: Doctoral Dissertation, Journal Article, Masters Thesis, Systematic Review **Search modes** - Boolean/Phrase	**View Results** (100) **View Details** **Edit**

3: **ERIC** (via EBSCO) 15th July 2015

Search ID#	Search Terms	Search Options	Actions
S14	S12 AND S13	**Search modes** - Boolean/Phrase	**View Results** (40) **View Details** **Edit**
S13	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9	**Search modes** - Boolean/Phrase	**View Results** (6,036) **View Details** **Edit**
S12	S10 AND S11	**Search modes** - Boolean/Phrase	**View Results** (1,122) **View Details** **Edit**
S11	medic* OR surg* OR medical education OR medical training	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (13,351) **View Details** **Edit**
S10	postgraduate* OR post-graduate OR resident* OR trainee* OR junior*	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (15,749) **View Details** **Edit**
S9	OSATS OR objective structured assessment of technical skills	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All),	**View Results** (1) **View Details** **Edit**

		Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	
S8	TX case based discussion OR TX case-based discussion	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (33) **View Details** **Edit**
S7	TX mini-PAT OR TX mini peer assessment tool	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (2) **View Details** **Edit**
S6	TX direct* observ* of procedural skills OR TX DOPS OR TX technical skills	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (733) **View Details** **Edit**
S5	TX mini-CEX OR mini clinical evaluation exercise	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (8) **View Details** **Edit**
S4	TX assessment N2 learning	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (2,499) **View Details** **Edit**

S3	TX formative feedback OR TX formative assessment OR TX formative evaluation	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (2,979) **View Details** **Edit**
S2	(multisource OR multi-source OR 360-degree OR 360 degree) N2 (feedback OR assessment* OR evaluation*)	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (84) **View Details** **Edit**
S1	((workplace-based N2 (assessment* OR feedback OR evaluation*)) OR (workplace based N2 (assessment* OR feedback OR evaluation*)) OR (work-based N2 (assessment* OR feedback OR evaluation*)) OR (work based N2 (assessment* OR feedback OR evaluation*)))	**Limiters** - Date Published: 19950101-20150731; Journal or Document: Journal Articles (EJ); Publication Type: Dissertations/Theses (All), Journal Articles, Reports - Research **Search modes** - Boolean/Phrase	**View Results** (92) **View Details** **Edit**

4. **PsycINFO** (via EBSCO Host) July 15th 2015

Search ID#	Search Terms	Search Options	Actions
S14	S12 AND S13	**Search modes** - Boolean/Phrase	**View Results** (3,591) **View Details** **Edit**
S13	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9	**Search modes** - Boolean/Phrase	**View Results** (152,663) **View Details** **Edit**
S12	S10 AND S11	**Search modes** - Boolean/Phrase	**View Results** (24,018) **View Details** **Edit**
S11	TX doctor OR TX medic* OR TX surg* OR TX medical education OR TX medical training	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (508,790) **View Details** **Edit**
S10	TX postgraduate* OR TX post-graduate* OR TX resident* OR TX trainee* OR TX junior*	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (55,131) **View Details** **Edit**
S9	TX OSATS OR TX objective structured assessment of technical skills	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (32) **View Details** **Edit**

S8	TX case-based discussion OR TX case based discussion	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (100) **View Details** **Edit**
S7	TX mini-PAT OR TX Mini peer assessment tool	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (7) **View Details** **Edit**
S6	TX direct observ* of procedural skills OR TX DOPS	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (57) **View Details** **Edit**
S5	TX assessment N2 learning	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (1,894) **View Details** **Edit**
S4	TX mini-CEX OR TX mini clinical evaluation exercise	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (52) **View Details** **Edit**
S3	TX formative feedback OR TX formative assessment* OR TX formative evaluation*	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (1,389) **View Details** **Edit**
S2	(multisource OR multi-source OR 360-degree OR 360 degree) N2 (assessment* OR feedback OR evaluation*)	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human	**View Results** (150,420) **View Details** **Edit**

		Search modes - Boolean/Phrase	
S1	(workplace-based N2 (assessment* OR feedback OR evaluation*)) OR (workplace based N2 (assessment* OR feedback OR evaluation*)) OR (work-based N2 (assessment* OR feedback OR evaluation*)) OR (work based N2 (assessment* OR feedback OR evaluation*))	**Limiters** - Published Date: 19950101-20150731; Publication Type: All Journals; Language: Dutch, English, French, German; Population Group: Human **Search modes** - Boolean/Phrase	**View Results** (150,549) **View Details** **Edit**

5. **British Education Index (**via EBSCO Host) 15th July 2015

<u>Search ID#</u>	Search Terms	Search Options	Actions
S14	S12 AND S13	**Search modes** - Boolean/Phrase	**View Results** (92) **View Details** **Edit**
S13	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9	**Search modes** - Boolean/Phrase	**View Results** (1,530) **View Details** **Edit**
S12	S10 AND S11	**Search modes** - Boolean/Phrase	**View Results** (1,824) **View Details** **Edit**
S11	TX medic* OR surg* OR medical education OR medical training	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (7,656) **View Details** **Edit**
S10	TX postgraduate* OR post-graduate* OR resident* OR trainee* OR junior* OR physician* OR doctor*	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (5,199) **View Details** **Edit**
S9	TX OSATS OR TX Objective structured assessment of technical skills	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (3) **View Details** **Edit**
S8	TX case based discussion OR TX case-based discussion	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (12) **View Details** **Edit**

S7	TX mini-PAT OR TX Mini peer assessment tool	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (3) **View Details** **Edit**
S6	TX direct* observ* procedural skills OR TX DOPS	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (5) **View Details** **Edit**
S5	TX mini-CEX OR TX Mini clinical evaluation exercise	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (27) **View Details** **Edit**
S4	TX assessment N2 learning	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (772) **View Details** **Edit**
S3	TX formative feedback OR TX formative assessment* OR TX formative evaluation*	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (858) **View Details** **Edit**
S2	(multisource OR multi-source OR 360-degree OR 360 degree) N2 (assessment* OR feedback OR evaluation*)	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (42) **View Details** **Edit**
S1	(workplace-based N2 (assessment* OR feedback OR evaluation*)) OR (workplace based N2 (assessment* OR feedback OR evaluation*)) OR (work-based N2 (assessment* OR feedback OR evaluation*)) OR (work based N2 (assessment* OR feedback OR evaluation*))	**Limiters** - Publication Date: 19950101-20150731; Publication Type: Academic Journal; Language: French, German, English **Search modes** - Boolean/Phrase	**View Results** (79) **View Details** **Edit**

6. **EMBASE** (Elsevier Platform) July 17th 2015

#22	#13 AND #16 AND #20 AND ([article]/lim OR [article in press]/lim OR [conference abstract]/lim OR [conference paper]/lim) AND ([dutch]/lim OR [english]/lim OR [french]/lim OR [german]/lim) AND [humans]/lim AND [1995-2015]/py	110
#21	#13 AND #16 AND #20	15
#20	#18 OR #19	119,124
#19	doctor* AND in AND training OR junior AND doctor*	22,942
#18	postgraduate* OR 'post graduate' OR resident* OR residency AND education	99,464
#16	#14 AND #15	13,999
#15	feedback OR assessment* OR evaluation*	3,097,357
#14	'clinical competence'/exp	44,672
#13	#3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12	6,763
#12	'objective structured assessment of technical skills' OR osats	294
#11	'direct observation of procedural skills' OR dops	623
#10	'case based discussion'	187
#9	mini AND peer AND assessment AND tool OR 'mini pat'	21
#8	'mini cex' OR 'mini clinical evaluation exercise'	178
#7	assessment NEXT/2 learning	659
#6	formative AND assessment* OR formative AND feedback OR formative AND evaluation*	1,946
#5	multisource OR 'multi source' OR '360 degree' OR 360 AND degree AND (feedback OR assessment* OR evaluation*)	2,864
#4	'work based' AND assessment* OR 'work based' AND evaluation* OR 'work based' AND feedback	75
#3	'workplace based' AND assessment* OR 'workplace based' AND evaluation* OR 'workplace based' AND feedback	141

7. **Science Direct** 22nd July 2015

pub-date > 1994 and (**workplace-based assessment OR formative assessment OR mini-CEX OR case-based discussion OR OSATS OR DOPS OR multisource feedback OR 360 degree feedback**) AND (**postgraduate trainee* OR postgraduate medical training OR medical education**)
[*Journals(- All Sciences -)*]   359

8. **Australian Education Index** (ProQuest Platform) 22nd July 2013

Set	Search	Databases	Results	Actions
S1	(workplace-based assessment) OR (formative assessment) **Databases:** • Australian Education Index	Australian Education Index	**651°**	Actions
S2	(direct observation of procedural skills) OR DOPS **Databases:** • Australian Education Index	Australian Education Index	**1°**	Actions
S4	mini clinical evaluation exercise **Databases:** • Australian Education Index	Australian Education Index	**1°**	Actions
S5	objective structured assessment of technical skills **Databases:** • Australian Education Index	Australian Education Index	**4°**	Actions
S6	case based discussion **Databases:** • Australian Education Index	Australian Education Index	**721°**	Actions
S7	mini peer assessment tool **Databases:** • Australian Education Index	Australian Education Index	**1°**	Actions
S11	(assessment for learning) AND (medical education) **Databases:**	Australian Education Index	**230°**	Actions

Set	Search	Databases	Results	Actions
	• Australian Education Index			
S12	S1 OR S2 OR S4 OR S5 OR S6 OR S7 **Databases:** • Australian Education Index	Australian Education Index	**1365°**	Actions
S13	SU.EXACT.EXPLODE("Graduate medical education") OR SU.EXACT("Medical education") **Databases:** • Australian Education Index	Australian Education Index	**1008°**	Actions
S14	S12 AND S13 **Databases:** • Australian Education Index	Australian Education Index	**39°**	Actions
S15	S14 OR S11 **Databases:** • Australian Education Index	Australian Education Index	**246°**	Actions
S16	S14 OR S11Limits applied **Databases:** • Australian Education Index **Narrowed by:** Entered date: 01/ 01/ 1995 - 22/ 07/ 2015	Australian Education Index	**221°**	Actions

9a: **Cochrane Database**

Individual searches for: Results: 25
      1. workplace-based assessment
      2. Mini-clinical evaluation exercise
      3. Mini-CEX
      4. Objective Structured Assessment of Technical Skills
      5. OSATS
      6. Case-based discussion
      7. Mini peer assessment tool
      8. Mini-PAT
      9. Multisource feedback
      10. 360 degree feedback

9b: **BEME published reviews**   Results: 0
All published reviews (www.bemecollaboration.org) were searched. The list of included articles of two reviews were hand-searched for relevant articles

Total citations = 7066

Appendix 5   Kirkpatrick's framework of educational outcomes

(with modifications by Barr et al (2000) and Steinert et al (2006))

Level I	Learner's reactions
Level 2a	Modification of attitudes and skills
Level 2b	Acquisition of knowledge and skills
Level 3a	Self-reported change in behaviour
Level 3b	Observed change in behaviour
Level 4a	Change in organisational practice
Level 4b	Benefits to clients or patients

Appendix 6   Modified BEME Review Coding Sheet

Can workplace-based assessment be used to identify and remediate poor performance among postgraduate medical trainees?

1. Administrative

Reference Number_____   Date: _____   Reviewer ☐ AB

☐RG

☐TH

Citation type:

☐ Journal article   ☐Non-peer review article   ☐ Official publication

☐ Book   ☐Thesis   ☐Other

Citation information

Title:

Author(s):

Publication: Year: Volume: Issue:
        Pages:

Search Method

☐Electronic search   ☐ Grey Literature   ☐Other

_____

2. Evaluation

a) *Aim of study*: ☐ Stated ☐ Not available

Aim/objectives:

b) *Conceptual framework*: ☐ Stated ☐ Not available

Conceptual/theoretical framework used:

c) *Research Design*

☐ Experimental Design ☐ Observational Study

  ☐ Randomised controlled trial   ☐ Case study/case series

  ☐ Pre-test/Post-test   ☐ Cross-sectional study

  ☐ Single group, no comparison   ☐ Cohort study

  ☐ Single group, repeated measures

☐ Qualitative Study ☐ Mixed Methods Study

  ☐ Grounded theory study

  ☐ Phenomenological

  ☐ Phenomenographic study

  ☐ Ethnographic

  ☐ Critical discourse analysis

  ☐ Narrative inquiry

Additional design comments/notes:

d) *Data collection methods* (tick all that apply)

☐ Interview   ☐ Focus Group   ☐ Observation   ☐ Questionnaire/survey   ☐ Trainee records/assessments

3. Study Context

Country/training jurisdiction:

Setting: ☐ Clinical setting ☐ Simulated setting ☐ Other (describe)

Population

      Trainee discipline (e.g. general medicine, ophthalmology)

      Year in Training (e.g. FY2)

Mode of recruitment (e.g. convenience or randomised sample; email invitation)

If targeted group (e.g. poorly performing) how was this group identified?

Sample size: Unit of analysis (if relevant):

Any additional contextual/descriptive data:

4. 'Intervention' details

WBA tool(s) (Tick all that apply)

☐ DOPS   ☐ Case-based Discussion   ☐ Mini-CEX   ☐ OSATS

☐ Mini-PAT   ☐ 360º/MSF   ☐ Other

_____

Any notable modifications to content or usual WBA process

```
[                                                                    ]
```

WBA Method used

☐ Trainee-led/initiated WBAs   ☐ Trainer-led/initiated WBAs

☐ Routine use WBAs   ☐ Targeted use

If routine use, were there guidelines/requirements for minimum numbers to be completed?

```
[                                                                    ]
```

If targeted use, how was target group identified (e.g. where poor performance suspected)?

```
[                                                                    ]
```

☐ Single WBA event   ☐ Multiple WBA events

Frequency of use of WBAs:

☐ weekly   ☐ monthly   ☐ annually   ☐ Other_____

5. Methodological Quality (modified BEME Quality Indicators Buckley et al, 2009)

QUALITY INDICATOR		YES/NO/ NOT APPLICABLE (N/A)
Research Question	Is the research question or hypothesis clearly stated?	
Study Subjects	Is the subject group appropriate for the study being carried out?	
Data Collection Methods	Are the methods used appropriate for the research question and context?	
Completeness of data	Attrition rates/acceptable questionnaire response rates?	
Risk of bias assessment	Is a statement of author positionality and a risk of bias assessment included?	
Analysis of results	Are the statistical and other methods of results analysis used appropriate?	
Conclusions	Is it clear that the data justify the conclusions drawn?	
Reproducibility	Could the study be repeated by other researchers?	
Prospective	Is the study prospective?	
Ethical Issues	Are all ethical issues articulated and managed appropriately? (was Ethical approval documented?)	
Triangulation	Were results supported by data from more than one source?	

6. Outcomes

a)   Educational outcomes

Kirkpatrick's framework of educational outcomes (*modifications by Steinert et al, 2012)

☐ Level 1   Learner reactions

☐ Level 2a   Modification of attitudes and skills

☐ Level 2b   Acquisition of knowledge and skills

☐ Level 3a*   Self-reported change in behaviour

☐ Level 3b*   Observed change in behaviour

☐ Level 4a   Change in organisational practice

☐ Level 4b   Benefits to clients or patients

b)   Trainee outcomes

☐ Progression/non-progression

☐ Remediation outcomes

☐ Performance changes on re-assessment (improved, dis-improved, no change)

c)   Practice-level outcomes (e.g. changes in use of WBA from routine to targeted)

d) *System-level outcomes* (e.g. modification of tools, introduction of new tools, discontinuation of tools)

Appendix 7   Weighted kappa calculation

Total screened by RG and AB = 836 (13.6%)

	Relevant	Unsure
AB	Wood Brown Bullock Burford	Sheehan Palmer Archer Francis Beckman Vivekananda-Schmidt Overeem Baker Berk Asemoto
RG	Wood Baker Brown Lin Burford	Berk Overeem Asemoto Vivekananda-Schmidt

Cohen's weighted kappa calculation

		RG			
		Relevant	Unsure	Irrelevant	Total
AB	Relevant	3	0	1	4
	Unsure	1	3	5	9
	Irrelevant	1	1	821	823
		5	4	827	836

Kappa = 0.587

Weighted kappa = 0.641

Weighted kappa (online calculation)

Appendix 8   Voting spreadsheet

	180	306	330	356	389	717	1174	4317	4610	4804	5172	6018	6179	6753	6811	6935
	Sheehan	Vivekanda	Palmer	Archer	Francis	Beckman	Coverdale	Aram	Weller	Benenson	Dowson	Mitchell	Pasquina	Wall	Whitehouse, Walzman	Liu, Pin
RG																
AB																
TH																
YS																
AS																

Appendix 9   Methodological quality evaluation (BEME indicators)

	1. Research question	2. Study subjects	3.Data collection methods	4.Complete-ness of Data	5. Risk of bias ax	6. Analysis of results	7. Conclusions	8. Reproducibility	9. Prospective	10. Ethical Issues	11. Triangulation	MQ 'score'
Archer et al, 2011	Yes	Yes	Yes	Yes	No[a]	Yes	No[b]	Yes	Unclear[c]	unclear[d]	Yes	7
Black et al, 2009	Yes	Yes	Yes	Unclear[e]	No[a]	Yes	Unclear[f]	Yes	No[g]	Unclear[h]	No[i]	5
Brown et al, 2008	Yes	Yes	No[j]	Yes	No[a]	Yes	Yes	Yes	No[k]	Unclear[d]	No[i]	6
Bullock et al, 2009	Yes	Yes	Yes	Yes	No[a]	Yes	Yes	Yes	No[g]	Unclear[d]	No[i]	7
Burford et al, 2010	Yes	Yes	Yes	Yes	No[a]	Yes	Yes	Yes	No[k]	Unclear[d]	No[i]	7
Chipp et al, 2011	Yes	Yes	Yes	Yes	No[a]	Yes	No[l]	Yes	Unclear[c]	Unclear[h]	Yes	7
Donato et al, 2015	Yes	Yes	Yes	Yes	No[a]	Yes	Unclear[f]	Yes	Unclear[c]	Unclear[d]	Yes	7
Ellul et al, 2012	Yes	Yes	Yes	Yes	No[a]	Yes	Unclear[f]	Yes	No[g]	Unclear[d]	No[i]	6
Hassell et al, 2012	Yes	Yes	Yes	Unclear[m]	No[a]	Yes	Yes	Yes	Yes	Unclear[h]	No[i]	7
Hesketh et al, 2005	Yes	Yes	Yes	Yes	No[a]	Yes	Yes	Yes	No[g]	Unclear[h]	No[i]	7
Hiemstra et al, 2011	Yes	Yes	Yes	Yes	No[a]	Yes	Yes	Yes	Unclear[c]	Unclear[h]	No[i]	7
Mitchell et al, 2011	Yes	Yes	No[n]	Yes	No[a]	Unclear[o]	Unclear[p]	Yes	No[g]	Unclear[d]	No[i]	4
Vivekananda-Schmidt et al, 2013	Yes	Yes	Yes	Yes	No[a]	Unclear[q]	Yes	Yes	No[g]	Unclear[d]	No[i]	6
Warm et al, 2010	Yes	Yes	Yes	Yes	No[a]	Yes	Unclear[r]	Yes	Unclear[c]	Unclear[d]	Yes	7
Whitehouse et al, 2007	Yes	Yes	Yes	Yes	No[a]	No[s]	Unclear[f]	Yes	Yes	Unclear[d]	Yes	7

	1	2	3	4	5	6	7	8	9	10	11	
Wood et al, 2006	Yes	Yes	Unclear[t]	Yes	No[a]	Yes	Yes	Yes	Unclear[c]	Unclear[h]	No[i]	6
Yao et al, 2000	Yes	Yes	No[j]	Yes	No[a]	Yes	Yes	Yes	No[k]	Unclear[h]	No[i]	6
Young et al, 2011	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear[h]	Yes	9

Rating key

Rating key	Questions
[a] No assessment of risk of bias [b] study design biased as MSF assessors already knew that trainee had been identified by others as underperforming – no attempt to determine this impact on assessor rating therefore conclusions limited [c] Timeline of study (vs intervention) unclear [d] no statement of informed consent [e] outcomes only provided for entire cohort of underperforming trainees; no data on individual sub-groups i.e. trainees identified using mini-PAT; also no data on which types of underperformance each method identified. No explicit statement of the denominator used in the study [f] limitations of study design not reflected in conclusions [g] retrospective [h] ethical approval and informed consent not described [i] no additional data used for triangulation of findings [j] Study reliant on recall, no documentation analysed [k] cross sectional study [l] conclusion claims not supported by small size of the study [m] issues with missing data [n] no information on pre-study identification of underperforming trainees [o] use of summary means explained but potentially limits impact of findings on ability of tool to identify dips in performance [p] comments on limitations of scores in identifying performance issues but no discussion of value of narrative feedback [q] no explicit framework used to guide interpretation/categorisation of narrative	1. Is the research question or hypothesis clearly stated? 2. Is the subject group appropriate for the study being carried out? 3. Are the methods used appropriate for the research question and context? 4. Are attrition rates/ questionnaire response rates acceptable ? 5. Is a statement of author positionality and a risk of bias assessment included? 6. Are the statistical and other methods of results analysis used appropriate? 7. Is it clear that the data justify the conclusions drawn? 8. Could the study be repeated by other researchers? 9. Is the study prospective? 10. Are all ethical issues articulated and managed appropriately? 11. Were results supported by data from more than one source?

feedback (e.g. content analysis, critical discourse analysis, grounded theory) [r] conclusion claims can be transferred, but not generalisable out of specific context [s] very limited analysis of qualitative interview data [t] no details on no of assessments performed per trainee; no information on progress-remediation outcomes where underperformance identified	

Appendix 10   Methodological quality of qualitative studies (CASP)

Authors	CASP Questionnaire									
	1	2	3	4	5	6	7	8	9	10
Sargeant et al, 2011	Yes	Yes	Yes	Yes	Unclear[a]	Unclear[b]	Yes	Yes	Yes	Clear
Weller et al, 2009	Yes	Yes	Yes	Unclear[c]	Yes	Unclear[d]	Unclear[e]	Yes	Yes	Clear

	CASP Key
[a] No topic guide provided with article and data saturation not discussed [b] Influence of the researchers on the data collection and analysis process not considered in the manuscript [c] No information provided on those who didn't attend. Method of recruitment lacks clarity. [d] Influence of the researchers on the data collection and analysis process not considered in the manuscript [e] informed consent not explicitly stated	1. Was there a clear statement of the aims of the research? 2. Is a qualitative methodology appropriate? 3. Was the research design appropriate to address the aims of research? 4. Was the recruitment strategy appropriate to the aims of the research? 5. Were the data collected in a way that addressed the research issue? 6. Has the relationship between researcher and participants been adequately considered? 7. Have ethical considerations been taken into consideration? 8. Was the data analysis sufficiently rigorous? 9. Is there a clear statement of findings? 10. How valuable is the research?

CHAPTER 5 FIGURES

Fig 1   Flow diagram of search and selection process