

THE BEME COLLABORATION
BEST EVIDENCE MEDICAL EDUCATION

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

Proposed Review Protocol; November 2014

CONTENTS

REVIEW TEAM..... 3

BACKGROUND TO THE TOPIC..... 5

REVIEW QUESTION, OBJECTIVES AND KEY WORDS..... 8

SEARCH SOURCES AND STRATEGIES 10

STUDY SELECTION CRITERIA..... 13

PROCEDURE FOR EXTRACTING DATA 15

SYNTHESIS OF EXTRACTED EVIDENCE..... 17

PROJECT TIMETABLE 18

CONFLICT OF INTEREST STATEMENT 19

PLANS FOR UPDATING THE REVIEW 19

REFERENCES 20

REVIEW TEAM

Aileen Barrett, Education Specialist, Royal College of Physicians of Ireland and PhD student, University College Cork, Ireland.

Aileen currently works as an education specialist in postgraduate medical training and is responsible for the implementation of research-informed changes to medical education and training processes along with responsibility for the design and delivery of faculty development workshops and initiatives including those focused on the implementation of workplace-based assessment. This work will form a significant component of her PhD work, which aims to explore the quality, value and effectiveness of workplace-based assessment in postgraduate medical education

Rose Galvin, PhD. Programme Leader HRB Centre for Primary Care Research, Royal College of Surgeons in Ireland. She has published a number of systematic reviews in both clinical and education-related topics

Yvonne Steinert, PhD. Director of the Centre for Medical Education and Richard and Sylvia Cruess Chair in Medical Education, Faculty of Medicine, McGill University, Montreal, Canada. Dr Steinert is also the Chair of the McGill BEME International Collaborating Centre (BICC) and has previously published two systematic reviews on faculty development under the auspices of BEME. She has also recently edited a book on Faculty Development in the Health Professions: A Focus on Research and Practice

Albert Scherpbier, MD, PhD. Dean, Faculty of Health, Medicine and Life Sciences, University of Maastricht, The Netherlands. Professor Scherpbier is also vice Chair of the Maastricht Medical University Center and has authored or co-authored a number of reviews, including BEME systematic reviews. He is also the lead reviewer's PhD co-supervisor

Ann O'Shaughnessy, PhD. Head of Education and Professional Development, Royal College of Physicians of Ireland. Her research experience includes a clinical systematic review

Mary Horgan, MD. Dean, School of Medicine, University College, Cork. Professor of Medicine and Consultant Physician in Infectious Diseases and is the lead reviewer's PhD supervisor

Tanya Horsley, PhD. Associate Director, Research Unit, Royal College of Physicians and Surgeons of Canada. Dr. Horsley currently holds an academic appointment in the Department of Epidemiology and Community Medicine, Faculty of Medicine, at the University of Ottawa. She is currently the Chair, Research Committee, for the Society of Academic Continuing Medical Education (SACME) and Assistant Editor, Journal of Continuing Education in the Health Professions (JCEHP). She has been involved with nearly 25 systematic reviews and is focused on improving methodological approaches to reviews of medical education

Collectively the group has extensive experience in systematic reviews, both in clinical and medical education research and in international collaborations, qualitative research methodologies - including phenomenology and grounded theory data collection and analysis methods - and in medical education. The lead for the review (AB) is currently undertaking a PhD in medical education.

Correspondence

Aileen Barrett
Education Specialist
Education and Professional Development
Royal College of Physicians of Ireland
Frederick House
19 South Frederick St
Dublin 2
Ireland
aileenbarrett@rcpi.ie
Phone: +353 87 9188 782

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 (Day, Grosso et al. 1990, Daelmans, Hoogenboom et al. 2004). In 1995 the first workplace-based assessment specifically designed to structure feedback following an observation of a clinical interaction – the Mini-Clinical Evaluation Exercise (Mini-CEX) was developed (Norcini, Blank et al. 1995) using research emerging theory from the UK and elsewhere on assessment-for-learning in which the goal of the interaction is to provide feedback on performance, informing a learning plan or action, and 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 (e.g. the recently developed McGill P-MEX tool to assess professionalism). As of 2012, there were 52 WBA ‘tools’ in use (Kogan et al, 2012).

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 (Bindal, Wall et al. 2011, Kogan, Conforti et al. 2012). Literature and anecdotal evidence suggests that trainers feel uncomfortable giving negative feedback and structuring learning plans for trainees (Kogan, Conforti et al. 2012) and that trainees view WBA as a ‘tick-box exercise’ (Bindal, Wall et al. 2011) with little impact on their learning and development. More recently, Watling, Driessen et al. (2012) 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 (2009) 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 (Mitchell, Bhat et al. 2013). The researchers did not find strong associations between trainees in difficulty and the level of complexity of the clinical cases they used for their WBAs 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 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 day varies so significantly in terms of methods and quality.

- Kogan, Holmboe et al. (2009) 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 (2010) 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, Wollersheim et al. (2009) 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, Salleh et al. (2012) reviewed 15 prospective studies of workplace-based assessments in postgraduate medical education. They too were unable to ascertain a definite link between those 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, Kramer et al. (2011), 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 assessment of performance and in high stakes situations (e.g in making progress judgments), these tools have yet to be validated fully.

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 in an unpublished internal review, 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 performing trainees.

An initial literature search has identified a number of studies looking at the identification of poor performance using specific tools (Black and Welch 2009, Mitchell, Bhat et al. 2013); 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 the inclusion of workplace-based assessment can assist in identifying poor performance, under what conditions and what effect, if any, using WBAs may have on remediating and changing clinical practice.

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:

- 'Poor performance' 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' (Mitchell, Bhat et al. 2013). 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 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 within Cleland et al., (2010). 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

1. To determine if the use of workplace based assessment(s) can be reliably used to identify poor performance among postgraduate medical trainees
2. To determine if workplace-based assessments can be used to effectively remediate poor performance among postgraduate medical trainees

Secondary

1. To determine the conditions under which the use of WBA is most useful 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 evaluated using Kirkpatrick’s framework of educational outcomes using Barr’s adaptations (Barr, Freeth et al. 2000) for medical education research. This model uses the following levels of evaluation (Table 1) We will also use the modification of level 3 (change in behaviour) proposed by Steinert, Naismith et al. (2012) to distinguish between self-reported and observed changes

Table 1: Kirkpatrick’s framework of educational outcomes

Level 1	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 is that it uses a hierarchical approach to evaluating evidence (Yardley and Dornan 2012), 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. 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
- BEME published reviews, Cochrane, DARE
- PsycINFO
- Science Direct

Our searches will be limited to 1995 to the most recent search date given the first 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

Sample MEDLINE search

Database:

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present.

Search Date: Nov. 4, 2014.

Retrieval: 1956 records.

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

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

1. Review of reference lists of included studies and review articles
(<http://www.ncbi.nlm.nih.gov/pubmed/21833989>)
2. 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
3. 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.
4. We will also conduct a citation search on Web of Science looking for studies citing any 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

	INCLUSION CRITERIA	EXCLUSION CRITERIA
Population	<ul style="list-style-type: none"> • Postgraduate medical trainees • Postgraduate surgical trainees 	<ul style="list-style-type: none"> • 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
Intervention	<ul style="list-style-type: none"> • Workplace-based assessments e.g. Mini-CEX, DOPS in direct observation situations (see Types of Interventions, pg 13) 	
Outcomes	<ul style="list-style-type: none"> • Studies that describe/report outcomes related to identification/remediation (see Types of Outcomes pg 14) 	
Research Design	<p>Studies which provide primary data for any of the outcomes above including, but not limited to, the following designs:</p> <ul style="list-style-type: none"> • Experimental and/or observational studies • Randomised and non-randomised studies • Prospective or retrospective cohort studies • Qualitative • Descriptive 	<ul style="list-style-type: none"> • Studies that do not report an outcome 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 WBAs 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 WPA 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 remedying 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' Reliability/validity data stated
- Descriptive / Contextual data
- Risk of bias assessment
 - 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
 - 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
 - Practice-level
 - Changes from routine to targeted (and vice versa) use of WBA
 - 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 (Buckley, Coleman et al. 2009) as this has been used in previous WBA systematic reviews (Miller and Archer 2010) 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 2)

Table 2: BEME Quality Indicators (Buckley et al, 2009)

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 subset 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 (2012) 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 (Cook and West 2012) will be undertaken.

In the case of qualitative studies included for analysis, we will use a qualitative meta-synthesis analysis method to explore the common themes and concepts emerging from the research studies (Walsh, Galvin et al. 2014)

PROJECT TIMETABLE

	PRE-REVIEW WORK JAN-OCT 2014	OCT	NOV	DEC	JAN 2015	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT-DEC 2015
Title registration, protocol development and review														
Database search and pilot coding sheet														
Coding of studies and data analysis														
Data synthesis														
Preparation of paper for publication														

CONFLICT OF INTEREST STATEMENT

The members of the research team do not report any conflict of interest

PLANS FOR UPDATING THE REVIEW

In keeping with the BEME systematic review requirements, the review will be updated within three years of completion. However, should the volume of published literature increase before that date, we plan to update the review as and when appropriate to ensure that it remains up to date and relevant.

REFERENCES

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

Bindal, T., D. Wall and H. M. Goodyear (2011). "Trainee doctors' views on workplace-based assessments: Are they just a tick box exercise?" Medical Teacher **33**(11): 919-927.

Black, D. and J. Welch (2009). "The under-performing trainee – concerns and challenges for medical educators." The Clinical Teacher **6**(2): 79-82.

Buckley, S., J. Coleman, I. Davison, K. S. Khan, J. Zamora, S. Malick, D. Morley, D. Pollard, T. Ashcroft, C. Popovic and J. Sayers (2009). "The educational effects of portfolios on undergraduate student learning: A Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11." Medical Teacher **31**(4): 282-298.

Cook, D. A. and C. P. West (2012). "Conducting systematic reviews in medical education: a stepwise approach." Medical Education **46**(10): 943-952.

Daelmans, H. E. M., R. J. I. Hoogenboom, A. J. M. Donker, A. J. J. A. Scherpbier, C. D. A. Stehouwer and C. P. M. van der Vleuten (2004). "Effectiveness of clinical rotations as a learning environment for achieving competences." Medical Teacher **26**(4): 305-312.

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

Kogan, J. R., L. N. Conforti, E. C. Bernabeo, S. J. Durning, K. E. Hauer and E. S. Holmboe (2012). "Faculty staff perceptions of feedback to residents after direct observation of clinical skills." Medical Education **46**(2): 201-215.

Kogan, J. R., E. S. Holmboe and K. E. Hauer (2009). "Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees." JAMA: The Journal of the American Medical Association **302**(12): 1316-1326.

Miller, A. and J. Archer (2010). "Impact of workplace based assessment on doctors' education and performance: a systematic review." BMJ **341**.

Mitchell, C., S. Bhat, A. Herbert and P. Baker (2013). "Workplace-based assessments in Foundation Programme training: do trainees in difficulty use them differently?" Medical Education **47**(3): 292-300.

Norcini, J. J., L. L. Blank, G. K. Arnold and H. R. Kimball (1995). "The Mini-CEX (Clinical Evaluation Exercise): A Preliminary Investigation." Annals of Internal Medicine **123**(10): 795-799.

Overeem, K., H. Wollersheim, E. Driessen, K. Lombarts, G. Van De Ven, R. Grol and O. Arah (2009). "Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study." Medical Education **43**(9): 874-882.

Pelgrim, E. A. M., A. W. M. Kramer, H. G. A. Mokkink, L. den Elsen, R. P. T. M. Grol and C. P. M. Vleuten (2011). "In-training assessment using direct observation of single-patient encounters: a literature review." Advances in Health Sciences Education **16**(1): 131-142.

Saedon, H., S. Salleh, A. Balakrishnan, C. Imray and M. Saedon (2012). "The role of feedback in improving the effectiveness of workplace based assessments: a systematic review." BMC Medical Education **12**(1): 25.

Steinert, Y., L. Naismith and K. Mann (2012). "Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19." Medical Teacher **34**(6): 483-503.

Walsh, M. E., R. Galvin, C. Loughnane, C. Macey and N. F. Horgan (2014). "Factors associated with community reintegration in the first year after stroke: a qualitative meta-synthesis." Disability and Rehabilitation **0**(0): 1-10.

Watling, C., E. Driessen, C. P. M. van der Vleuten, M. Vanstone and L. Lingard (2012). "Understanding responses to feedback: the potential and limitations of regulatory focus theory." Medical Education **46**(6): 593-603.

Yardley, S. and T. Dornan (2012). "Kirkpatrick's levels and education 'evidence'." Medical Education **46**(1): 97-106.

