Which is the most effective assessment tool to measure the ability to self-reflect and respond with insight? A Best Evidence In Medical Education Systematic Review.

BEME Protocol 2015 v1.3 revised April 2016

**Group Leader**
Dr Julie C. Williams  
School of Oral and Dental Sciences  
University of Bristol  
Lower Maudlin Street  
Bristol  
BS1 2LY  
Email: julie.williams@bristol.ac.uk  
Phone: 07872395326

**1 Title**
BEME (1) Protocol: Which is the most effective assessment tool to measure the ability to self-reflect and respond with insight?

**Group members and their areas of relevant expertise and interest**

**Dr Julie Williams** BDS, MA (Ethics of Healthcare), MO orth, DDS, DPDS, MGDP is an Academic Clinical Lecturer in Orthodontics at the University of Bristol School of Oral and Dental Sciences. She was previously a Foundation Trainer of graduate dentists for 9 years and has co-written a Guide for Assessment in Dental Education. This is her second BEME review, her first being Professional (non-technical) competencies of graduate veterinarians (completed 2015, BEME Review No.38) and she is currently undertaking a PhD in Medical Education. She will act as the PI for this study.

**Professor Sarah Baillie** BVSc, MSc, PhD, PFHEA, MRCVS holds a Chair in Veterinary Education at University of Bristol and has led the recent major curriculum review of the veterinary programme. She is involved in teaching various aspects of professionalism and her research interests include simulation within Medicine and Veterinary Medicine and assessment. She is on the executive committee for the curriculum review within Medicine. She has been involved in two previous BEME reviews: Assessing competence in veterinary education: where’s the evidence? (2008) and Professional (non-technical) competencies of graduate veterinarians (completed 2015, BEME Review No.38).

**Cath Borwick** MSc is an Information Scientist and systematic reviewer for the School of Social and Community Medicine, University of Bristol and the Research Design Service South West, NIHR. She has been the information scientist for several published reviews, responsible for aspects of searching the literature, developing data extraction forms and
contributing to the analysis and writing up of the review and associated publications. She teaches a range of research skills, academic and digital literacies for the Medical undergraduate programme and this is her first BEME review.

Dr Martin Cake BSc BVMS PhD is an Associate Professor in Veterinary Anatomy in the College of Veterinary Medicine, Murdoch University, Australia. He has a strong interest in professional competencies from roles as both an educator (as contributor to Murdoch University’s Veterinary Professional Life modules) and in curriculum development. He is currently leading an international project researching employability and resilience in the veterinary context. This is his second BEME review, the first investigating evidence for the importance of professional skills for graduate veterinarians (2015 BEME Review No.38).

Dr David Dymock BSc, PhD is a Senior Teaching Fellow at the University of Bristol School of Oral and Dental Sciences where he is Education Director. He is also Education Director for undergraduate programmes, including medical, dental, veterinary science and intercalation programmes, within the University of Bristol Faculty of Health Sciences and is Co-Director of the Centre for Health Sciences Education. He has previously chaired the Oral Microbiology and Immunology Group of the British Society of Oral and Dental Research, and was conferred an Excellence in Dental Education award from the Association of Dental Education in Europe in 2012. This is his first BEME review.

Dr Ellayne Fowler BA, MA, EdD, PGCE, FHEA is a teaching fellow on a postgraduate medical education Masters where she leads the dissertation stage. As an educationalist with a background in qualitative research she is involved in curriculum design and supporting educational research across the Health Sciences Faculty. This is her first BEME review.

Professor Anthony Ireland BDS, MSc, PhD, FDS RCS, M.Orth RCS, PG Cert Dent Ed, FHEA Prof Ireland is Training Programme Director for the South West of England Doctor of Dental Surgery (DDS) programme in Orthodontics and was previously the Training Programme Director in Orthodontics for the Bristol University Distance Learning for Dentists programme. He is also Graduate Director for the School of Oral and Dental Sciences, University of Bristol and responsible for both the postgraduate taught and research programmes in the School. In this role he is involved in curriculum design and implementation. Prof Ireland teaches orthodontics at all levels including therapist, dental undergraduate, foundation dentist and postgraduate dentist. This is his first BEME review.

Sheena Warman BSc, BVMS, DSAM, DipECVIM-CA, SFHEA, MRCVS is a Senior Clinical Fellow in Small Animal Medicine at the University of Bristol Veterinary School. She is deputy programme director for the veterinary programme, and chairs the Assessment and Feedback group within the Centre for Health Sciences Education. She contributes to staff development for Medical, Dental and Veterinary professional programmes across the
Faculty of Health Sciences. She is currently enrolled in an EdD programme at the University of Bristol; her thesis is exploring how reflection is enacted by recent veterinary graduates. This is her first BEME review.

**Funding source**
Nil
Abstract

**Background:** Health professionals, including doctors, nurses, dentists and veterinary surgeons are expected to have the capacity to reflect upon their practice throughout their careers and to demonstrate insight by responding appropriately to self-reflection. These abilities, including an understanding of personal strengths and weaknesses, are necessary for continued professional development, especially if there is evidence of sub-optimal performance (2). Indeed the ability to self-reflect with insight is referred to within the medical, dental, veterinary and nursing undergraduate curricula as a core, or key skill (3-5). It is therefore important to nurture and grow such skills within all undergraduate professionals. To facilitate this, a robust and responsive means of measurement is required that can be repeated throughout a course in order to evaluate baseline levels and subsequent development during professional training.

To date there is no such universally accepted measurement tool. This systematic review will address the question of which tools are available to measure self-reflection and insight within either the workplace or adult education and summarise the evidence available to support the use of these tools. It is envisaged that a suite of suitable tools could eventually be gainfully employed during the education and development of the undergraduate healthcare professional.

**Methods:** Using BEME (Best Evidence in Medical Education) Collaboration (1) review guidelines, electronic databases including PsycINFO, Medline, ERIC and CINAHL will be searched for studies where an instrument or tool is used to measure the ability to self-reflect in an adult educational or work-place setting. Data from each included study such as the sample, the methods and the results of any additional observation of self-reflection or insightful response will be extracted by 2 independent reviewers and entered into a modified BEME coding sheet, using Excel. The reviewers will also appraise the quality of the studies using the BEME coding form “Strength of Findings” model, adapted from a previous BEME review (6) in order to identify the strength of evidence for the utility of the assessment instrument, such as its reliability, validity, educational impact, practicality and cost.

**Results:** It is expected that the studies will be heterogeneous and a mix of qualitative and quantitative study designs. The output of these results will be presented in tabular form with a list of the tools used to evaluate the ability to self-reflect and respond with insight. Of particular interest will be those studies with repeated measures of these abilities within the same sample and those where several measures of these abilities have been used at the same time to permit triangulation of the findings.

**Discussion:** Studies within such diverse fields as business, education, engineering, healthcare, law and management describe the need to train their personnel to develop the
ability to self-reflect and to respond with insight. The first step in evaluating the impact of these programmes is to be able to measure these skills and attributes over a period of time. This systematic review will inform the development of a suitable tool which could eventually be used locally, nationally and potentially internationally to assess progress within this important area during undergraduate professional programmes.

**Systematic review registration:** The topic was successfully registered with BEME 17th August 2015

**Keywords:** Self-reflection, measurement, instrument, tool, education or work, undergraduate or postgraduate, self-awareness, insight, assessment.
2 Background (including rationale and aims of review)

Why are we interested?
As tutors of undergraduate healthcare professionals, we have become aware of the greater emphasis placed upon self-reflection both nationally and locally within our own School of Healthcare Sciences at Bristol. The four national regulatory bodies, with responsibility for the content of professional programmes, namely the General Medical Council, the General Dental Council, the Nursing and Midwifery Council and the Royal College of Veterinary Surgeons, refer to critical reflection and its importance within their recommended standards for education (4, 5, 7, 8). Two of these bodies also state that the new graduate should either be able to self-reflect or to develop the capacity for self-audit (5, 8). Self-reflection is not clearly defined in the literature although a study to assess student reflection in medical practice defines reflection in the context of medical professionalism as

“the conscious weighing and integrating of views from different perspectives, a necessary prerequisite for the development of a balanced professional identity” (p67) (9).

In addition, professionals must have the capacity to change when presented with evidence of sub-optimal performance and for this insight is a necessary part of the process (2). Hays (2) refers to the insight of a clinician as being part of judgment, an understanding of oneself as a professional and how to behave in a similar manner to an expert professional in the workplace. This resonates with a more recent definition of insight by Brown et al. (10):

“how and why I, and those I interact with, behave, think and feel as we do and for me to adapt my behaviour accordingly” (10).

If self-reflection and insight are considered to be essential for the fledgling clinician, we as tutors and educators need to know how to recognise, nurture and grow these traits or skills within undergraduate professionals.

Testing knowledge, skills or performance, either to acknowledge competence or to identify areas for improvement, requires a robust means of measurement and on the whole this is relatively straightforward to perform for example within dental education (11). What is less clear is the way we can define and measure self-reflective capacity or ability. A previous systematic review (12), undertaken to create a blueprint of methods to assess professionalism in its entirety, acknowledged this limitation, but suggested both self-reflection and insight could be assessed, possibly by adapting tools already in use. However no such tools were clearly identified.

This led us to the following questions: How should we define the terms self-reflection and insight, in the context of healthcare education? Then, reiterating the two questions posed by Sumssion and Fleet (13) two decades ago, why should we and how can we assess self-reflection and insight?
How do we define self-reflection and insight?
The term insight, in the context of healthcare, as defined by Brown et al. (10) implies that a clinician has an understanding of what others, such as patients whom they treat or colleagues within the wider professional community, expect of a professional within the workplace. However, the terms reflection, critical reflection, self-reflection and self-regulation are poorly defined in the healthcare literature (14) and this may be because they signify complex concepts (15). Authors of a recent review have proposed a five-component model of reflection from which evolved their definition, namely

“Reflection is the process of engaging the self in attentive, critical, exploratory and iterative interactions with one’s thoughts and actions, and their underlying conceptual frame, with a view to changing them and with a view on the change itself” (16).

This has considerable strengths; it clearly unifies several models of reflection to create a definition that is simple to understand yet has sufficient depth to convey the complexity of the construct. It is interesting that the proposed interaction with one’s thoughts or actions is only with “a view to change”. Justification of a view would also require reflection even if the intention of the professional were to maintain that view. In the related field of critical thinking, Facione (17) summarises an American consensus from a Delphi process to define critical thinking:

“purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological or contextual considerations upon which the judgment is based” (17).

These definitions overlap. Both describe a journey of self-discovery, underpinned by a framework of previous knowledge and experience. There are two key differences. Firstly the definition of reflection acknowledges that the process involves one’s own thoughts and actions as the subject under scrutiny, whilst critical thinking is not so specific; it may be a self-regulated judgment about a text, an event, an experiment or perhaps another’s actions. Secondly the definition of critical thinking does not assume that the thinker is planning to change their view(s), simply they should be able to interpret, analyse, evaluate and infer from the judgment in context. There is no such freedom within the definition of reflection by Nguyen et al. (16), which is surprising since the reflector may decide, perhaps for themselves or in discussion with their peers or mentor, that their thought or action was the correct one in the circumstances.

Nguyen et al. (16) are also keen to emphasise that reflection is more than “the sum of its parts” yet the figure they have used to illustrate their definition (p1181)(16) appears to use the components as a series of building blocks, related to either content or process on the way to increasingly reflective practice. Previous work by Grant et al. (18) which aimed to develop assessment tools to measure self-reflection, has focused on dividing the process into component parts with the implication that this is somehow easier to grasp than the
whole. Grant et al. (18) proposed four stages- an understanding of the need for self-reflection, the willingness to practise (or engagement), the ability to conduct self-reflection and finally whether the student demonstrates accurate self-reflection, often termed insight. It is assumed that the student will progress through each stage although the same authors state

“One may spend considerable time in self-reflection without gaining insight”

(p824)(18).

Whilst Grant’s model is not described as a cycle, a more traditional description of reflection as a cyclical process was proposed by Koole et al. (19). These authors identified inconsistent definitions of terms used in the field of reflection, both nationally and internationally, as one of four factors considered to confound the assessment of reflection. They created an eclectic model of the reflective process, informed by the work of Dewey, Boud, Schön, Kolb, Moon and Mezirow, which describes reflection in three phases; reviewing the experience, critical analysis and reflective outcome. Each phase is divided into two operational indicators, for example ‘reviewing the experience’ is indicated by ‘the ability to describe an event adequately’ and ‘the ability to identify and describe essential elements and one’s own thoughts and feelings’. ‘Critical analysis’ is indicated by ‘the ability to ask searching questions’ and ‘to answer searching questions within a defined frame of reference’. Finally the ‘reflective outcome’ relies on the ‘ability to draw conclusions’ and ‘to describe concrete learning goals and plans for future action’ (19). Whilst this model does not use the term self-reflection per se, the reflection process could be applicable to reflection on one’s own experiences as self-reflection or about the experiences of others. This eclectic model of reflection was therefore identified as a practical conceptual framework for our review of the literature.

**Why do we undertake assessment?**
Assessment of the undergraduate within professional disciplines has three purposes. Firstly assessment, especially with effective feedback, is known to drive learning (20), secondly assessment allows tutors to monitor what students have learnt and thirdly it provides information to those who evaluate the undergraduate clinician against the required standard of competency, prior to certification. Whilst these would appear to be noble aims, the assessment methods used should also ideally be able to discriminate for excellence and also encourage appropriate learning within the students. If the stakeholders consider the tool to be suitable, this could increase student motivation to improve in this area. For a skill such as self-reflection, this could help mitigate the anecdotal perception amongst the professions and Faculty that students are ‘fed up’ with the process of reflecting or even discussing the process of reflecting, but show limited ability to reflect or to respond with insight.
Without a suitable form of assessment, it is difficult to know if any perceived improvement in a student’s ability is due to an educational intervention, to a clinical episode, to something that has happened to the student within their personal life or to a timely discussion with a peer, role model or mentor. Careful testing of progress over time with an assessment tool could help to clarify if, how and when students develop the required self-reflective skill or ability and could also help to identify any part of the course that has a negative impact on this ability. Schuwirth and van der Vleuten’s general overview of the theories used in assessment (21) describes the benefits of assessment-for-learning in diagnosing areas for educational development of the student. This can then lead to the provision of tailor-made solutions for each student to allow them to achieve competency or even excellence in the long term. A programmatic approach to assessment with multiple opportunities for expert judgment of the undergraduate is recommended (21).

**Why should we assess self-reflection and insight?**

A non-systematic critical narrative review by Ng et al. (15) was undertaken in response to concerns surrounding a reductionist approach to the teaching and assessment of reflection within medical education. Three dominant trends in the contemporary use of reflection within medical education were presented as;

- a means to help learners achieve a particular outcome,
- a means to improve the self and
- reflection within assessment, as a competency in itself or as a means to achieve other competencies.

Ng et al. concluded that treating reflection as an epistemology of practice, rather than in the forms outlined above, permits broad and deep understanding of the field within medical education (15). The conclusion does not expand on how this could be applied in the short term, but it is important to acknowledge the authors’ approach, since it firmly places reflection as a worthwhile global theory to be desired for its own worth, not just as a means to an end.

Other authors (22) have raised similar concerns about the “ritualizing” of reflection with the practice being a recipe to follow, “without reference to context or outcomes” (22). Whilst these comments are not about the assessment of the ability to self-reflect, but rather about the process of reflection itself, it would be expected that the application of the paraphernalia of rubrics, questionnaires and portfolios to measure self-reflective ability could also add to the risk of ritualisation and the process of self-reflection being taken out of context.

**How can we assess the ability to self-reflect and respond with insight?**

Four factors have been identified as confounders for the assessment of reflection (19). These are:
• inconsistent definitions of terms, as outlined previously,
• lack of agreed standards to interpret the results of assessments,
• threats to the validity of current methods of assessing reflection and
• the influence of internal and external contextual factors on the assessment of reflection. Examples were provided such as motivation for the task, or the presence of meaningful relationships with the facilitator or tutor as part of the reflective process which could also be considered threats to the validity of assessment methods.

Agreed standards imply a global understanding of just how self-reflective a competent practitioner needs to be, although this is not well defined within educational standards of professional regulatory bodies (4, 5, 7, 8). As with many constructs particularly those included within professionalism, it may be considered easier for clinicians or peers to identify an absence of these abilities and insight, for example as part of disciplinary hearings after a breach of professional care. Doctors who require disciplinary action in the United States are more likely to have demonstrated unprofessional behaviour whilst in medical school (23), particularly severe irresponsibility and severely diminished capacity for self-improvement. The method of assessing capacity for self-improvement per se was not detailed in this study but negative excerpts about professional behaviour were extracted from documents in the doctor’s files dated before graduation, including the Deans’ letters of recommendation to residency programmes. This identifies faculty opinion as one method of setting professional standards, although these students were still successfully certified fit for practice, despite recorded concerns.

A small-scale study (24) to assess the accuracy of peer-assessment, self-assessment and expert assessment of a graduate group task, in the business sector rather than a clinical setting, found substantial convergence between expert and peer rankings. There was however little convergence between a graduate ranking themselves in the group and expert/peer ranking suggesting that self-assessment may not always be accurate, compared to the view of a group of experts or peers. This is important because the continued registration of a clinician following a professional misdemeanour relies on judgment by their peers and lay-people as part of professional regulation. Insight failure has been shown to be an early sign of a doctor-in-difficulty, demonstrated by a rejection of constructive criticism often associated with defensiveness and counter-challenge (p85, 25).

Two scales have been developed to help individuals to judge their own ability to self-reflect, which is rather paradoxical since it relies on a certain initial level of self-awareness but circumvents indirect observation by another (19). The Groningen Reflection Ability Scale or GRAS (25) was devised to measure personal reflection in medical practice and education whilst the Self-Reflection and Insight Scale (SRIS) is a tool initially developed by Grant et al. (26) which has recently been validated with medical students (27). Both scales use 5-point Likert scales and the inherent problem of self-completion of these is acknowledged by the authors of GRAS (25), although they consider self-judgment on personal characteristics is
not automatically less accurate than peer judgments, citing work by Hofstee et al. (1998) to support this statement.

Returning to Nguyen et al’s review, it is recommended that

“For assessing reflection, we should not measure the final destination of reflection or the distance travelled using it, but, rather, the reflective journey itself” (16).

If either the component parts of a process or the overall ability to self-reflect and show insight is to be assessed, then a suitable tool would need to be practical or in educational terms, to have utility. The definition of the term utility used to inform this review is:

Utility of an assessment = reliability x validity x educational impact x acceptability x cost (28)

An initial search of the literature conducted in March 2015 failed to find a systematic review of the evidence for the utility of instruments to measure the ability to self-reflect and respond with insight. However, methods used within medical education to teach the skills required to self-reflect are frequently reported and include the use of a portfolio (29), reflective writing (30), journals (31) and role modeling of reflective practice (32). A BEME review is planned to look at the use of methods of teaching self-reflective skills for undergraduate medical students (33). The assessment of learning following teaching using these methods remains unclear. For example the portfolio, either electronic or written, has received considerable attention within the medical education literature as a means to provide evidence of student reflection (6). Ker’s reflective ability assessment instrument (34) has been used in one study (35) of final year medical students to mark a self-reflective portfolio of evidence regarding infection control measures, although there is no consensus regarding how portfolios are used or assessed, with or without the specific measurement of reflective ability or insight.

Alternatively rubrics have been used to assess reflective writings, for example Wald et al. (36) uses “systematized feedback frameworks” as part of mentor-enhanced reflection and development of professional identity. This includes an assessment rubric, the ‘Reflection Evaluation for Learners’ Enhanced Competencies Tool (REFLECT)’ which is primarily used to teach the “art of reflection” and “accompany learners at critical junctures on the educational journey” (p698, 36). Both of these methods require a degree of interpretation of the event leading to the reflection and adequate description of the event requires written or oral communication skills which may unduly influence the validity of assessment of the reflective process (19). Other threats to the validity of the process could include the reason for undertaking the assessment (formative, summative or certification), where the assessment is undertaken (clinical or non-clinical settings) and the relationship between the assessor and the student.
During our review of the literature, it became clear that assessment of self-reflection has been undertaken within many disciplines including education (37), management within the workplace (38), musicianship (39), leadership (40) and many professions other than healthcare, including the ministry (41), teaching (42), accountancy (43) and business (44). These studies frequently described the use of tools such as portfolios or self-reported measures although, as with medical education, there were limited examples of repeated measures of self-reflections and insight. One study within one medical school (27) using the SRIS, showed students to have poor correlation of self-reflective ability and academic scores, although there was a slight increase in the recorded score for the need for self-reflection within one cohort of students following a year of clinical training. Unfortunately the same cohort had a slight reduction in score for engagement in self-reflection and insight over the same period. The authors suggested that this could be explained by the time-pressures of the final year of the course, which may require considerable time for memorising facts for the examinations, rather than spending time in periods of self-reflection.

The importance of developing healthcare professionals who have the ability to self-reflect and respond with insight is therefore supported within the literature, both within medical education and also in the wider context of other professional disciplines. This systematic review will address the question of which tools are available to measure these attributes or skills within either the workplace or adult education so that a suite of suitable tools could eventually be gainfully employed during the development of the undergraduate healthcare professional. This assessment process is expected to be cyclical or even helical and in a similar manner to clinical audit, could lead on to a second cycle of review, analysis and response which could all be subsequently assessed ideally with repeated use of the same tool(s).

3 Review topic question and objectives
Which is the most effective assessment tool to measure the ability to self-reflect and respond with insight?

Primary objectives of this review
To determine which tools have previously been used in adults within a workplace or educational setting to measure self-reflection and insight.
To explore, analyse and synthesise the evidence related to the utility of those tools identified.
**Additional questions:**
With any identified tools, what is the evidence for their
- Reliability
- Validity
- Cost effectiveness
- Educational impact
- Practicality

Also, against what other criteria are the results measured? Is there any evidence for use of these tools in the context of undergraduate healthcare professional education and specifically for repeated testing using these tools?

### 4 Study selection criteria

The review group discussed the findings of the literature review and agreed the conceptual framework for the review. The review question and study selection criteria were developed using the SPICE format (45). The opportunity to explore non-healthcare workplace settings, in addition to educational settings, was discussed and it was felt that other fields might have developed a useful tool that could be adapted for use during the education and training of healthcare professionals. It was decided to exclude clinical studies of patients where the patient’s ability to self-reflect is evaluated as part of their clinical treatment, for example in psychiatry or chronic clinical conditions such as diabetes. The population to be studied, adults older than 16 years of age, was selected since this evidence is more likely to be practically applied to the testing of undergraduates rather than tools used during primary (elementary) or secondary education. An inclusive approach to intervention was undertaken to explore the breadth of the literature with two specific exclusions; studies with self-reported self-reflective abilities in the absence of a validated questionnaire (40) and studies which referred to assessment of insight within a specific but different context such as the Intercultural Development Inventory (46) or clinical reasoning (47) were excluded. Studies where several tools have been used concurrently will be of particular interest, such as a self-administered questionnaire with a validated scale in addition to observed reflective behaviour.

The following inclusion and exclusion criteria were agreed (Table 1):
<table>
<thead>
<tr>
<th>SPICE (45)</th>
<th>INCLUSION CRITERIA</th>
<th>EXCLUSION CRITERIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>Educational or workplace setting</td>
<td>Clinical studies of a patient’s self-reflection or self-awareness e.g. weight loss, management of diabetes, self-harm programmes, mental health issues</td>
</tr>
<tr>
<td>Population</td>
<td>Young adults Students Undergraduates or postgraduates Adults</td>
<td>Children &lt; 16 years of age</td>
</tr>
<tr>
<td>Intervention</td>
<td>Instrument for the assessment of self-reflection or self-awareness for example (but not exclusively) scale, tool, measure, questionnaire Studies with self-completed scales that were scored by computer or another person</td>
<td>Studies with solely student self-reported ability or improvement in ability to self-reflect or respond with insight unless part of a validated externally scored questionnaire Tests of clinical reasoning Tests of cultural awareness</td>
</tr>
<tr>
<td>Comparison</td>
<td>Other measurement or observation of non-technical values or skills such as observed ability to self-reflect or respond with insight Multiple measurement of the same sample using one instrument</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>Systematic reviews, controlled trials (randomised and non-randomised), quasi-experimental designs such as repeated measures or pretest/post-test, cohort studies including cross-sectional studies, case control studies, qualitative studies, mixed methods studies and theses that include detail of an assessment tool and some detail of its validity, reliability, cost-effectiveness, practicality and/or educational impact.</td>
<td>Descriptive studies outlining benefits of e.g. a training intervention with no method of the assessment process.</td>
</tr>
</tbody>
</table>

Table 1 Development of Inclusion and Exclusion Criteria
5 Search sources and strategies

a) Scoping review

A scoping search was performed in June 2015 by one reviewer (JCW) and a subject librarian (CB) with previous experience of conducting systematic reviews. This helped to identify proposed search terms within the literature including education, teaching, train*, university or college or further education or school, in-service train*, work*, personnel management, staff development, management, educational measurement (including assessment), tool or Instrument or scale or test, self-reflect*, self-aware*, self-regulat*

* = truncation

The group leader (JCW) and the subject librarian (CB) developed the scoping search strategy initially by iterative aggregation of these search terms, with subject headings where thesauri exist, in conjunction with free text using truncation and appropriate Boolean operators. The term insight was found to produce an unworkable number of false hits in Ovid Medline and Embase since many authors include the term in their conclusion, notwithstanding the subject or topic under examination. The term was however successfully used in combination with the term self-reflection for four databases, namely the Australian Education Index, British Education Index, CINAHL and ERIC. The sample scoping search strategy as used for one database is described below:

Sample Ovid MEDLINE search

Database: PubMed

Search date 8th June 2015, filtered by Specific Year Range 1975-2014

Retrieval: 1870

Search strategy:

1 Education/
2 (undergraduate* or postgraduate* or student*).tw.
3 education$.tw.
4 Teaching/
5 train*.tw.
6 (universit* or college* or further education* or school*).tw.
7 Inservice Training/
8 work*.tw.
9 teach*.tw.
10 exp Personnel Management
11 Staff Development/
12 staff development.tw.
13 management.tw.
14 exp Educational Measurement/
15 (tool*or instrument*or scale* or test*).tw.
Of the 1870 studies retrieved, 1130 had been published in the previous 5 years.

b) Process of sifting
A process for sifting of studies by title and abstract was developed (Table 2).

Table 2 Process for sifting of studies by title and abstract

The sifting process (Table 2) was applied to the scoping search results in order to inform the final search strategy, using the first 100 studies backwards by date. 78/100 were excluded (0 duplicates, 2 English abstract only, 11 irrelevant for example reflective markers tested on fish farms, 7 studied young subjects, 27 studied patients, 30 did not measure self-reflection, 1 referred to cultural self-reflection).
A meeting of all the reviewers was held to discuss the findings of the scoping search. A random sample of 5 included and 5 excluded studies was distributed to the review team in order to help further refine the search terms for the definitive search strategy. The search terms were refined and adapted for example the term “self-regulation” was removed from the search string which reduced the numbers of irrelevant clinical studies with no loss of potentially relevant articles. After discussion there was agreement between JCW and the other reviewers regarding the inclusion and exclusion criteria and these were confirmed. Applying the filter for English-language only was considered a necessary restriction for the final search due to the wide scope of the search and the absence of funding for translation.

The search will cover a 40-year period (January 1975- December 2015). The search will include:

- electronic databases including Medline, Embase, CINAHL (Cumulative Index to Nursing and Allied Health Literature), ERIC (Education Resource Information Centre), British Education Index, Australian Education Index, PsycINFO, Web of Knowledge (Science Citation Index and Social Science Citation Index), Cochrane Library, The Philosopher’s Index, Research and Development Resource Base (RDRB), Campbell Collaboration
- ancestry searching using the Science Citation Index, to check for instruments or tools mentioned within citations
- contacting authors of studies where further planned studies using the tools are mentioned or where clarification of the detail of the study is required. Contact was made by email either from the contact details listed within their papers or following an Internet search for their current research institution. These authors were also asked if they were aware of any other unpublished (grey) literature that could be included in this review.
- grey literature such as the following web sites http://www.pdfsearchengine.org (e.g. 2.1 million hits first 10 pages searched and 6 papers included for full text analysis, http://www.crd.york.ac.uk, search.proquest.com

The primary database search will be supplemented by an updated electronic search in February 2016. The full search strategy will be included as an appendix in the final report.
6 Method of study selection

The search strategy described in section 5 will be used by one reviewer (JCW) to identify suitable studies and the exclusion process described in Table 2 will be applied, using the titles and abstracts of potential studies. The titles and abstracts of 35 studies (10% of the first tranche) will be independently considered for inclusion by a second member of the review group (AJI) and inter-rater reliability will be measured. Any dispute over which papers to include at the abstract stage will be decided by checking the full text. Studies selected will be imported to EndNote X7.01 reference management software (Thomson Reuters, Philadelphia). The full text of all potentially eligible papers will be obtained although a small number of studies may have to be excluded since the full text for all eligible papers may not be available for scrutiny. Details of this process will be reported as a flow diagram in the results section.

The full text of potentially relevant studies will then be read by the same reviewer (JCW) and examined for eligibility criteria. The authors will be contacted if there is a need for clarification of details of the study or to obtain the full text of the study, for example as unpublished theses. A final list of eligible studies to be included in the coding process will be collated and a series of folders have been constructed on a shared drive in order to distribute PDF documents for eligible papers and the coding sheet to each reviewer.

7 Procedure for extracting data

a) Development of coding sheet

Three members of the review group with previous experience of systematic reviews (SB, CB, JCW) constructed an initial coding sheet using studies from the literature search to inform the process. Three reviewers (SB, SW, JCW) piloted the coding sheet by independently coding 5 further papers from the scoping search. The results were discussed and the coding sheet was amended. The coding sheet was distributed electronically to all review group members and all reviewers coded the same study. A meeting was held to discuss the findings and minor changes were made to the coding sheet, for example a list of study designs taken from the BEME website (1) was attached to an additional spreadsheet for ease of reference. The final coding sheet for extracting data is attached as Appendix 1.

b) Extraction of data

Two members of the group will review each included paper independently using the agreed data extraction sheet (Appendix 1). The coding sheets will be completed electronically for assimilation by the group leader.
8 Appraisal of studies

There are 6 reviewers within the team. Each paper will be appraised independently by two reviewers once it has been deemed suitable for inclusion, using the BEME coding system shown in Table 3, as used in previous BEME reviews (6).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>No clear conclusions can be drawn. Not significant. LOW</td>
</tr>
<tr>
<td>2.</td>
<td>Results weak/ambiguous, but there appears to be a trend</td>
</tr>
<tr>
<td>3.</td>
<td>Conclusions can probably be based on the results</td>
</tr>
<tr>
<td>4.</td>
<td>Results are clear and very likely to be true</td>
</tr>
<tr>
<td>5.</td>
<td>Results are unequivocal. HIGH</td>
</tr>
</tbody>
</table>

Table 3 BEME coding system for strength of evidence (1)

In the case of disagreement during the appraisal of the studies, a third team member will be asked to independently review the paper in question and the three coders for those studies will discuss their findings in order to reach a consensus. Consistency between reviewers in terms of data extraction quality and quantity will be monitored by the group leader using a shared access spread sheet. Inter-rater agreement will be quantified by the Kappa statistic comparing global quality scores from the first two reviewers.

The reviewers will also be asked to give their opinion about the practicality of each tool described within the studies for the purpose of use as an assessment tool within the Faculty of Health Sciences at the University of Bristol.

Following data extraction of the first five papers for each reviewer, the group will meet again to discuss coding and any discrepancy in consistency will be discussed. In light of the heterogeneity of the papers retrieved during the scoping search, it is considered unlikely that a meta-analysis of the data will be possible, however this will be confirmed once all the papers have been retrieved.

9 Synthesis of extracted evidence

Data tables will be constructed in Microsoft Excel spreadsheets detailing study design, population characteristics, instruments used and the context of their use, the results and conclusions. From these tables we will identify which tools are being used and evaluate the evidence for their use in their current setting as previously entered on the coding sheet, with particular emphasis upon their reliability, validity, cost-effectiveness, practicality and educational impact (using a modified version of Kirkpatrick’s level of outcomes, Table 2). The reviewers will also record whether the coded level of educational impact of the
assessment tool or instrument was directly reported within the study or assigned by the review team member.

<table>
<thead>
<tr>
<th>Level</th>
<th>Reaction</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Reaction</td>
<td>Participants’ views on the learning experience, its organisation, presentation, content, teaching methods and quality of instruction</td>
</tr>
<tr>
<td>Level 2A</td>
<td>Learning-change in attitudes</td>
<td>Changes in the attitudes or perceptions among participant groups towards teaching and learning</td>
</tr>
<tr>
<td>Level 2B</td>
<td>Learning- modification of knowledge or skills</td>
<td>Knowledge- acquisition of concepts, procedures and principles. Skills- acquisition of thinking/problem-solving, psychomotor and social skills</td>
</tr>
<tr>
<td>Level 3</td>
<td>Behavioural change</td>
<td>Transfer of learning to the workplace or willingness of learners to apply new knowledge and skills</td>
</tr>
<tr>
<td>Level 4A</td>
<td>Results- change in the system/organisational practice</td>
<td>Wider changes in the organisation, attributable to the educational programme</td>
</tr>
<tr>
<td>Level 4B</td>
<td>Results- change among the participants’ students, residents or colleagues</td>
<td>Improvement in student or resident learning/performance as a direct result of the educational intervention</td>
</tr>
</tbody>
</table>

Table 4 Appraisal of Educational Impact, using modified Kirkpatrick Scale from BEME Collaboration specimen coding sheet (1)

10 Translation into practice

It is anticipated that a valid and reliable tool, or components of tools will be found from this review to measure the ability to self-reflect and respond with insight. This tool(s) or combination of tools could then be used in educational practice and, depending on the context, may need to be further validated. Potential applications would include:

- Measuring a baseline level of these skills when students embark upon their undergraduate training. It would be ideal to nurture and develop these baseline abilities.
- Monitoring individuals’ progress over time. Early identification of those undergraduate students who are struggling to reflect and demonstrate insight allows a programme of
measures to improve these skills to be implemented at an earlier stage, rather than toward the end of their clinical training or even afterwards. Regular review using the tool could also capture any unexpected decrease in these abilities during the course.

- Evaluating the impact of a particular educational intervention upon the student which has been designed to improve their skills, thereby creating individualised educational programmes that could focus upon developing insight and self-reflective ability as required.

The tool could be used locally for testing within the Faculty of Health Sciences, University of Bristol, but could be adopted more widely. Planned research using the tool includes:

- Comparing and contrasting levels of self-reflective ability between students on similar courses such as undergraduate veterinary, dental and medical programmes.
- Detecting any change in the ability to self-reflect and respond with insight by annual testing of the same cohort of students within each School.
11 Project timetable

<table>
<thead>
<tr>
<th>Task</th>
<th>Estimated time required</th>
<th>Proposed scheduling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refining the research question</td>
<td>Present</td>
<td>Completed</td>
</tr>
<tr>
<td>Scoping search and develop coding sheet</td>
<td>1-3 months</td>
<td>Completed by October 2015</td>
</tr>
<tr>
<td>Refining BEME coding sheet</td>
<td>2 months</td>
<td>Completed by mid December 2015</td>
</tr>
<tr>
<td>Literature search and retrieval of full text of potential studies</td>
<td>3-5 months</td>
<td>To be completed by April 2016</td>
</tr>
<tr>
<td>Data extraction and coding</td>
<td>2-4 months</td>
<td>To be completed by July 2016</td>
</tr>
<tr>
<td>Draft report</td>
<td></td>
<td>To be completed by September 2016</td>
</tr>
<tr>
<td>Final report</td>
<td></td>
<td>To be completed by November 2016</td>
</tr>
</tbody>
</table>

12 Conflict of interest for all review group members

The group members report no conflicts of interest

Plans for updating the review

The group plans to maintain an up-to-date bibliography related to the review question. From this any significant changes in the evidence base available would lead to an update of the review.

Changes to the protocol

Any changes to the protocol will need to be reviewed and agreed by BEME.
References

1 The BEME Collaboration 2016; Available online from: http://bemecollaboration.org. [Accessed on 12 February 2016]


5 Nursing and Midwifery Council Competencies for Entry to the Register. 2010; Available online from: https://www.nmc.org.uk/standards/code/N.M.C. [Accessed on 03 February 2016]


