

2017-07-31

# Towards an understanding of how appraisal of doctors produces its effects: a realist review

Brennan, Nicola

<http://hdl.handle.net/10026.1/9845>

---

10.1111/medu.13348

Medical Education

Wiley

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*

## **TOWARDS AN UNDERSTANDING OF HOW APPRAISAL OF DOCTORS PRODUCES ITS EFFECTS: A REALIST REVIEW**

Nicola Brennan, Marie Bryce, Mark Pearson, Geoff Wong, Chris Cooper, Julian Archer

Journal: Medical Education

### **ABSTRACT**

#### ***Introduction***

Revalidation was launched in the UK to provide assurances to the public that doctors are up-to-date and fit-to-practice. Appraisal is a fundamental component of revalidation. Approximately 150,000 doctors are appraised annually costing an estimated £97million over 10 years. There is little understanding of the theory of how and why appraisal is supposed to produce its effects. A realist review of the literature was utilised to explore these issues as they generate context-mechanism-outcome (CMO) configurations, resulting in the creation of theories of how and why appraisal of doctors produces its effects.

#### ***Methods***

A programme theory of appraisal was created by convening stakeholders in appraisal and searching a database of research on appraisal of doctors. Supplementary searches provided literature on theories identified in the programme theory. Relevant sections of texts relating to the programme theory were extracted from included articles, coded in NVivo and synthesised using realist logic of analysis. A classification tool categorised included articles contributions to programme theory.

#### ***Results***

125 articles were included. Three mechanisms were identified; dissonance, denial and self-affirmation. The dissonance mechanism is most likely to cause outcomes of reflection and insight. Important contexts for the dissonance mechanism include the appraiser being highly-skilled, the appraisee's working environment being supportive and the appraisee having the right attitude. The denial mechanism is more likely to be enacted if the opposite of these

contexts occur and could lead to game-playing behavior. A skilled appraiser was also important in triggering the self-affirmation mechanism resulting in reflection and insight. The contexts, mechanisms and outcomes identified were however limited by a lack of evidence to reliably further refine the CMO configurations.

### ***Conclusion***

This review makes a significant contribution to our understanding of appraisal by identifying different ways that appraisal of doctor's produces its effects. Further research will focus on testing the CMO configurations.

## **1. INTRODUCTION**

### ***1.1 Background***

Medical appraisal is an educational intervention traditionally based around a formative, developmental meeting between two professionals supported by information gathered from the full scope of practice.<sup>1</sup> It became a contractual requirement for all United Kingdom (UK) General Practitioners (GP) and Consultants in the National health Service (NHS) in the early 2000's,<sup>2,3</sup> with approximately 150,000 doctors in practice now appraised each year.<sup>4</sup> The appraisal process was originally intended to give feedback on a doctor's performance, map a doctor's progress and outline further development requirements<sup>5</sup> and was not designed as an assessment of competence that a doctor either passed or failed.<sup>6</sup>

With the introduction of revalidation by the General Medical Council (GMC) in 2012, the importance of medical appraisal was elevated. Revalidation is based on a doctor's participation in five annual appraisals and aims to provide objective assurance to the public that a doctor is up-to-date and fit-to-practise.<sup>7</sup> Revalidation necessitates the review of all doctors' licences to practise every five years by a responsible officer (RO).<sup>8</sup> The RO is typically the most high-ranking doctor in the organisation and they make their judgement using the appraisal output summary. If the ROs have any fears regarding a doctor's fitness to practise, they are legally obliged to refer doctors to the GMC for further investigation. The Department of Health estimated revalidation will cost £97million over 10 years.<sup>9</sup>

The interlinking of appraisal and revalidation, and the potential impact of revalidation on the appraisal process, has resulted in much debate. This has been compounded by disparate views on the goal of revalidation; either to detect 'bad apples', necessitating a summative approach and low criteria, or a system where all doctors improve; necessitating evolving standards and a developmental model.<sup>10</sup> However, developing a workforce is not the same as making sure it is safe and confusion over the dual purpose of revalidation may have an unintended impact on the educational value of appraisal.<sup>11</sup> Revalidation is purported to have impacted on the appraisal process in other ways including increased time costs<sup>12</sup>, collusion<sup>13</sup> and engagement in appraisal on a superficial level by disinterested appraisers and appraisees

(game-playing),<sup>14</sup> and might reduce any regulatory control that appraisal was intended to deliver.

Before we can fully understand the potential impact of revalidation on the appraisal process, it is important to understand how appraisal of doctors is intended to produce its effects i.e. what are the causal mechanisms at work in the appraisal process? Despite the increased importance of the appraisal process to the medical profession, and the change in nature of medical appraisal, there is little understanding of the theory of how and why appraisal is supposed to improve doctors' performance. Scallan et al's scoping review of the literature on appraisal found that to date the literature has mainly focused on: 'engagement' from the perspective of the doctor being appraised; exploration of a different models of appraisal (internal vs external appraisal, peer/practice, appraisal for revalidation); the benefits of appraisal; GPs' perceptions of the problematic link between appraisal and performance assessment or revalidation; administrative/management issues; and developing appraisers skills.<sup>15</sup>

Identifying the causal mechanisms in the appraisal process will contribute to the theory of how doctor's performance is improved by appraisal. It may then be possible to understand the impact of revalidation on the appraisal process. Hitherto no research of this type has been conducted. It will have also have practical implications, in that an increased understanding of 'how and why appraisal works' will aid decision-making around modifying and implementing appraisal processes at a local level. At an international level, the UK is unique in that certification of doctors relies heavily on appraisal.<sup>16</sup> There are some countries that do use a form of doctor appraisal<sup>17</sup> but this is not related to certification. However appraisal relies on a number of processes (e.g. reflection, feedback on performance, developing insight) that are commonly used when regulators and professional bodies in other jurisdictions assess the fitness to practice of doctors. What this paper will attempt to show is not only how and why these processes work in the UK context, but also point out where the findings will be 'externally valid' to other systems around the world.

## **2. METHODS**

### ***2.1 Research Questions***

The purpose of this review was to understand how and why appraisal of doctors produces its effects. The review questions included:

1. What are the mechanisms by which appraisal of doctors is believed to result in its intended outcomes?
2. What are the important contexts which determine whether the different mechanisms produce their intended outcome?
3. In what circumstances is appraisal likely to be effective?

We define an effective appraisal as one that results in a positive behaviour change in a doctor's practice that may ultimately benefit patients.

### ***2.2 Realist Review***

A realist review approach was used to address the review questions. Realist review is a theory orientated and explanatory approach to evidence synthesis. The literature is interrogated to develop and refine the theories that support the intervention being studied (in this review, appraisal) to explain what works, for whom, in what circumstances, in what respects.<sup>18</sup> A central part of a realist review is the development of a programme theory, which is an 'abstracted description' usually including a diagram that delineates the key functions, strategies or activities of an intervention, the intended outcomes of the intervention and the mechanisms that contribute to particular outcomes.<sup>19</sup> A realist synthesis takes a generative approach to causation – "to infer a causal outcome (O) between two events (x and y), one needs to understand the underlying mechanism (M) that connects them and the context (C) in which the relationship occurs".<sup>20</sup> Our review followed Pawson's five practical stages in conducting realist reviews.<sup>18</sup> A key feature of the realist review method is that it is iterative and the process frequently necessitates going back and forward between the different steps as the programme theory evolves.

### 2.2.1 Step 1: Locate Existing Theories

In order to develop an initial programme theory of appraisal we needed to locate existing theories. This was achieved in two ways. Firstly, we consulted with key stakeholders in appraisal including doctors/appraisees, appraisers, academic experts on appraisal, the GMC, ROs and human resources personnel. Formal meetings were held with the stakeholders at 3-4 monthly intervals throughout the review and took the form of facilitated discussions centered on evolving program theory. Stakeholders were identified through their involvement in previous research conducted by JA, MB and NB. Secondly, we searched the literature to find existing theories on how and why appraisal is supposed to work. These theories formed the foundation of the preliminary programme theory which was then tested with data from studies included in the review.

### 2.2.2 Step 2: Search Strategy

We undertook two searches. The search for data to develop, confirm, refute or refine ('test') aspects of the programme theory focused on well-established bibliographies generated from research on appraisal and revalidation conducted by the 'Collaboration for the Advancement of Medical Education Research and Assessment' (CAMERA) research team.<sup>21-26</sup> The CAMERA research team has conducted a program of research on appraisal and revalidation since 2009.

This search strategy was different to the strategy outlined in our protocol,<sup>27</sup> which involved searching electronic databases (e.g. EMBASE), but was adopted because our piloting of searches indicated that it was an effective way of locating the relevant literature. All references relating to appraisal and revalidation were compiled into a single library using Endnote X7.4 (Thomson Reuters Corporation, Toronto, Ontario) software programme. The library consisted of 463 references, including articles from peer-reviewed journals, books, reports and websites. The references of included papers were also searched for relevant articles. Searches of the CAMERA database were performed in April 2015 and updated in May 2016. The database was kept up-to-date. Members of the stakeholder group and review team were also canvassed to identify relevant literature.

The supplementary searches were purposive and undertaken when we identified that we needed more data on specific theories on different aspects of the programme theory on how

appraisal is supposed to work (e.g. how feedback on performance is meant to change behaviour, the process of developing insight, the relationship between insight and reflection, behaviour change theories, experiential learning theories, self-affirmation theories). Medline and Google Scholar were searched using topic keywords. Once relevant references were found, backwards and forwards citation searching techniques were used to identify further relevant papers. Relevant literature was also identified by members of the review team and stakeholder group.

### 2.2.3 Step 3: Study Selection Criteria and Procedures

Documents used to 'test' the programme theory were selected based on relevance (i.e. does the source contain evidence/data that we can use to develop and/or 'test' aspects of programme theory?). Documents were not excluded because of their type and we included editorials, opinion pieces, commentaries, process evaluations, qualitative research, programme manuals and systematic reviews. Using a preliminary set of inclusion/exclusion criteria the references in the database were then screened. The criteria were deliberately broad:

Inclusion criteria:

- Aspect of intervention/appraisal – all documents that contain information on the appraisal meeting between the appraisee and appraiser
- Study design – all study designs
- Types of settings – documents relating to healthcare settings
- Types of participants – all documents about doctors (in any specialty and at any career stage)
- Outcome measures – all appraisal related outcome measures.

Studies focusing only on the supporting information gathered by doctors for the appraisal meeting, the summary of the appraisal meeting, doctors' personal development plan (PDP) and the appraisers' statements or just revalidation were excluded. Studies about medical students were also excluded.

A randomly selected sample of 10% of the identified articles were evaluated by two review authors (NB and MB) using the inclusion/exclusion criteria. The remaining 90% were screened

by one reviewer (NB). Any discrepancies were discussed between NB and MB until agreement was reached. NB held the casting vote.

#### 2.2.4 Step 4: Extracting and organising data

A hybrid tool was used to classify the included articles with regards to how much they were likely to contribute to developing programme theory (Table 1). This existing tool categorised sources as conceptually rich (thick) or thin (weaker) in their explanation power.<sup>28-30</sup> Again, a random sample of 10% of the included articles were independently assessed by two reviewers (NB and MB) and any discrepancies were discussed until agreement was reached.

In a traditional systematic review data extraction is typically done using a standardised form however the realist review approach synthesises information through note-taking and annotation. Full texts of the included articles were imported into NVivo 10 (QSR International Pty Ltd, Doncaster, Vic, Australia). Sections of texts which we interpreted as relating to one or more aspects of the programme theory were then coded firstly by conceptual 'themes' and then by context, mechanism or outcome later during analyses.

#### 2.2.5 Step 5: Data Synthesis

The data coded in NVivo were synthesised using a realist logic of analysis. More details can be found in our review protocol.<sup>27</sup> Data from included documents were used to 'test' and refine each part of the preliminary programme theory. For each outcome identified by the programme theory, we searched for data to support inferences about the likely causal mechanisms and the contexts in which those mechanisms might be triggered. The final realist programme theory is explained using narrative synthesis, text and figures. Findings were reported according to the RAMESES publication standards for realist syntheses.<sup>19</sup>

### **3. RESULTS**

#### ***3.1 Summary of Studies***

Ninety-six articles focusing on appraisal of doctors were identified to develop and 'test' the initial programme theory (Figure 1). Of the 463 references identified in the CAMERA database, 394 were excluded based on title and abstract. Sixty-nine references were read in full with a further 27 citations identified for inclusion through citation tracking.

The 96 included articles were published in 30 different journals, with the majority being in *Education for Primary Care* and the *British Journal of General Practice* (Table 2). The majority of studies were carried out in the UK in Primary Care. The articles addressed a variety of topics in relation to appraisal. The majority of articles were opinion pieces or editorials. The majority were classified as having a thinner description.

For programme theory development, a further 29 papers were identified that contained theories that contributed to an understanding of how appraisal is supposed to work. The contents of these were synthesised which helped the development and refinement of different aspects of our initial programme theory.

A full list of the 125 articles are available from the corresponding author by request.

### **3.2 Programme theory development**

#### **3.2.1 From appraisal intervention strategy to realist programme theory**

The included articles were used to create an outline of the appraisal intervention (Figure 2). From this outline of the appraisal process the realist programme theory was iteratively developed.

Our programme theory (Figure 3) concentrates on disaggregating between Step 4 (the appraiser providing feedback) and Step 5 (inclusion of objective/action in PDP agreed) of the appraisal intervention strategy (Figure 2). In between these two steps we believe there are two potential intended intermediate outcomes i.e. reflection, insight<sup>31</sup> and then the unintended outcome of game-playing.<sup>32</sup> These intermediate outcomes may lead to the outcome of the appraisee including an objective in their PDP (Figure 3). There are also a number of further outcomes after this including the appraisee carries out what is in their PDP, followed by behaviour change, followed by improved performance. Our realist review specifically focuses on the outcomes of reflection, insight and game-playing.

### **3.3 Context-Mechanism-Outcome Configurations**

For the outcomes within our initial programme theory (Figure 3) we explored potential realist explanations for how different outcomes within the programme theory (reflection, insight and game-playing) have come about. Our exploratory realist explanations are expressed in the form of context-mechanism-outcome (CMO) configurations. In other words, for each outcome we propose which mechanism could have caused it and the contexts in which the mechanism is likely to be triggered.

However, a major caveat of this realist review was the limited rich data we were able to find in the included documents with which to develop, confirm, refute and refine the programme theory. While there was sufficient data in the literature to support the importance of certain contexts and outcomes, there was much more limited data on the relationships between them and on mechanisms. This meant that we were unable to fully and reliably configure the CMOs. Even where evidence was available, the thin nature of much of the evidence limited the depth of the explanations we could develop. The following examples of data extracts from the included articles demonstrate the differences between the rich and thin data. An example of data with a rich description was:

“The value of appraisals also depends on whether the appraiser is skilled [context] in conducting appraisals and is supportive, focused on the future, and participative. Badly conducted appraisals depress rather than improve performance [outcome].<sup>33</sup>

While an example of data with a thin description was:

“appraisal has a tendency to operate superficially within medicine, with doctors engaging in creative ‘game playing’ [outcome] toward its procedural requirements”.

34

Thus the CMO configurations we present in the following sections are supported by data where possible but had to be supplemented by the content expertise of the review team and stakeholder group.

#### **3.3.1 Dissonance**

This CMO configuration includes the mechanism dissonance which we inferred from the included sources to be necessary in order to achieve the desired appraisal outcomes of reflection and insight (Figure 3).<sup>22, 32, 35, 36 37 31 38, 39</sup> Cognitive dissonance is a well-recognised concept and refers to a situation where a person has conflicting attitudes/beliefs/behaviours. This produces a feeling of discomfort which leads to an alteration in the person's attitudes/beliefs/behaviours in order to reduce the discomfort and restore balance.

In medical appraisal, the concept of dissonance has the properties of a mechanism. The feedback from the appraiser is used as a strategy to create a situation or context that is likely to trigger dissonance<sup>40-45</sup> (mechanism) leading to reflection (an intermediate outcome), which may then eventually lead to an 'a-ha' moment<sup>46</sup> or insight (outcome). Accordingly, the appraisee will seek to resolve this dissonance in some way. They can either accept the data presented to them and re-evaluate their performance (insight / the 'a-ha' moment), or they can seek to resolve the dissonance in another way (for example when the denial mechanism is activated - see below).

The skills of the appraiser (context) were reported as being crucial to an effective appraisal.<sup>14, 33, 34, 47-52</sup> The success of appraisal partially depends on the quality of the feedback provided by the appraiser.<sup>50</sup> Thus if the appraiser is highly skilled (context) and able to effectively deliver feedback (context), this triggers the dissonance mechanism which is more likely to lead to the appraisee developing insight (outcome).

If the appraisee has the "right" characteristics and approach it in a proactive manner (context) the appraisal is more likely to be effective.<sup>52 53 54</sup> Incongruent feedback is more likely to trigger dissonance and result in reflection and/or increased insight (outcome) and ultimately behaviour change (outcome).<sup>44</sup> The same argument could be made for the context of time available to prepare for appraisal.

Our findings of the behaviour of the dissonance mechanism is consistent with Conlon's adaptation of Kolb's Learning cycle which depicts appraisal as a formalised way of guiding a professional through the learning cycle (Figure 4).<sup>38</sup> Reflection is the connection between experience and the generation of ideas, which then leads to behaviour change. Appraisal works on the reflection part of the learning cycle; by forcing appraisees to reflect and then providing feedback on the reflections and they then develop insight and change their

behaviour.

### 3.3.2 Denial

This CMO configuration includes the mechanism of denial<sup>40, 43</sup> which we inferred from the included data. It led to the less desirable outcome of game-playing or treating the appraisal process like a tick-box exercise (Figure 3).<sup>32, 34, 55, 56</sup> The denial mechanism interacts with the dissonance mechanism. One way of reducing cognitive dissonance is denial of the feedback or validity of the data being presented.<sup>43</sup> When an appraisee treats the appraisal like a tick box exercise<sup>32, 34, 55, 56</sup> they are essentially going through the process of appraisal without actually benefitting from the feedback intervention strategy. Besides changing behaviour, a doctor can reduce his or her commitment to the goal (context). Such a reduction would reduce dissonance by making the discrepant behaviour less relevant to the clinician and easier to deny.

We infer that there are several contexts in which the denial (mechanism) is likely to trigger game-playing or tick boxing behaviour (outcome). These contexts are the opposite to those that trigger the dissonance mechanism and include the appraiser being unskilled<sup>13</sup> the appraisee not having the right approach to appraisal and the appraisee not having enough time available to prepare for appraisal.<sup>14, 35, 57-59</sup>

Our findings of the behaviour of the denial mechanism are consistent with Payne and Hysongs physician feedback model i.e. and non-acceptance (or denial) of feedback tends to result in no behaviour modification.<sup>60</sup>

### 3.3.3 Self-Affirmation mechanism

This CMO configuration includes the mechanism of self-affirmation which we infer leads to the appraisee maintaining or improving performance (outcome) (Figure 3). There was some evidence to support this mechanism in the literature where appraisees reported the value of receiving confirmation that their performance was satisfactory, a recognition of their

achievements and identification of their strengths.<sup>12, 36 38, 50 58</sup> Self-affirmation is understood in psychology as an act that demonstrates one's adequacy based on the "premise that people are motivated to maintain the perceived worth and integrity of the self."<sup>61</sup> In any given day there are numerous events that are understood as relevant to the self in one way or another and this helps people to repeatedly recharge their sense of adequacy. A self-affirmation is a happening that validates a person's adequacy e.g. positive feedback on a skill.<sup>61</sup> Well-timed affirmations can improve performance, education, health and relationship outcomes.<sup>62</sup>

In medical appraisal, the concept of self-affirmation acts as a mechanism. If an appraisee's performance is strong in a particular area of clinical practice, a skilled appraiser (context) identifying these strengths (context) provides positive feedback confirming that they are performing well (context). This leads to self-affirmation (mechanism) of their clinical practice and to appraisees reflecting on their performance (outcome) thus maintaining or improving performance (outcome). This mechanism is particularly important when doctors are working in relative isolation.

Another context likely to influence whether the self-affirmation mechanism is triggered, relates to the appraiser and appraisee's relationship and in particular whether the appraisee respects and trusts the appraiser. If an appraiser is trusted and respected<sup>54</sup> by the appraisee (context) they are more likely to trust and appreciate the positive feedback they are given. If they do not trust and/or respect their appraiser (context) then it is likely that the feedback could lead to denial (mechanism) which leads to game-playing (outcome).

Our findings of the behaviour of the self-affirmation mechanism is congruent with Cohen's cycle of adaptive potential in which a positive feedback loop between the self and the social system can over time promote adaptive outcomes.<sup>62</sup> Path (a) in the diagram is particularly relevant i.e. as a result of being self-affirmed, the person achieves more adaptive outcomes e.g. better performance. Similarly part of Payne and Hysongs physician feedback model depicts a similar outcome i.e. when doctors receive favourable feedback they feel a sense of pride and try to ensure performance is maintained.<sup>60</sup>

#### **4. DISCUSSION**

#### **4.1 Statement of principal findings**

This realist review sought to explore the mechanisms and contextual factors in the appraisal of doctors. We developed a programme theory to explain how appraisal of doctors is supposed to produce its effects. This review included 125 articles which were extracted and synthesised to create CMO configurations centering on 3 mechanisms; (1) dissonance, (2) denial and (3) self-affirmation. We propose that it is through these mechanisms that appraisees reach both desirable and undesirable outcomes from appraisal. Contexts in which the *dissonance* mechanism is more likely to be triggered include the appraiser being highly-skilled, the appraisee's working environment being supportive of appraisal and the appraiser having the right attitude towards appraisal. The *dissonance* mechanism is most likely to result in reflection and insight which is more likely to lead to the appraisee changing their behaviour. The *denial* mechanism is more likely to be enacted if the opposite of these contexts occur i.e. the appraiser is unskilled, the doctor's environment is not supportive of appraisal, and the appraisee lacks the 'right' attitude towards appraisal. The denial mechanism is more likely to trigger the outcome of game-playing or treating appraisal like a tick-box exercise. Finally, if an appraisee's performance is strong in a particular area of clinical practice, a skilled appraiser identifying these strengths and providing positive feedback confirming that they are performing well, provides a self-affirmation which leads to the outcomes of reflection and insight. A trusting relationship between the appraisee and appraiser was also an important context for the self-affirmation mechanism and arguably for the dissonance mechanism too.

While we have identified 3 different mechanisms in the appraisal process, it would be reasonable to hypothesise that all three mechanisms – dissonance, denial and self-affirmation – could be present in one appraisal encounter in any combination. In an appraisal meeting a number of aspects relating to doctor's performance are discussed with the potential for each triggering a different mechanism and thus a different outcome. The structure of the programme theory allows for this complexity.

#### **4.2 Comparisons with existing literature**

This is the first realist review of appraisal of doctors and there has been no literature to date that has developed CMO configurations on this area. Thus we are cannot compare our proposed theory with existing literature due to the originality of our work. However, parts of the dissonance CMO configurations are consistent with theories of Kolb's learning cycle<sup>63</sup> and Williams et al's<sup>64</sup> stages of behaviour change i.e. participation in the appraisal process leads to reflection and the development of insight and ultimately to behaviour change. The components of the self-affirmation CMO configuration are consistent with Cohen et al.'s cycle of adaptive potential,<sup>62</sup> in that as a result of being self-affirmed a doctor's performance is improved. Some elements of the whole programme theory are consistent with Payne and Hysongs physician feedback model e.g. when doctors receive favourable feedback the feel a sense of pride and try to ensure performance is maintained and non-acceptance or denial results in no behaviour modification.<sup>60</sup>

#### ***4.3 Strengths and Limitations of the review***

The main strength of this review was the use of realist methodology. The knowledge created by this method has resulted in an empirical contribution to the existing body of literature, as previous research has not focused on identifying contexts or mechanisms that result in a successful appraisal. The included literature on the appraisal of doctors was exhaustive and was supplemented by literature from a variety of different areas to support our programme theory. While many of the concepts, like cognitive dissonance, self-affirmation, are not new what this realist review achieves is to bring disparate bodies of literature together linking them to appraisal. Furthermore, while this review focuses on an intervention implemented in the UK, this review is about a system that relies on a number of commonly used processes (e.g. feedback on performance) and thus our findings are 'externally valid' to other similar systems in place around the world. The review also has local practical implications, in that an increased understanding of 'how and why appraisal works' will aid decision-making around modifying and implementing appraisal processes at a local level.

In order to reduce errors in the screening, data extraction and quality assessment of sources, a proportion (10%) was carried out independently by two reviewers. The RAMESES quality standards for realist syntheses were followed throughout the review process. Input on

evolving programme theory and emerging findings was provided throughout the review by key stakeholders in appraisal.

The main limitation of this review relates to the lack of good-quality data in the included studies to effectively configure our CMO configurations. Thus, the explanatory power of this review was limited due to the absence of rich data. Empirical data on appraisal of doctors which contains elements of contexts and outcomes and can be used to make inferences about mechanisms is required. Literatures on other sectors in which appraisal has been implemented including nursing and higher education could have been explored however this was not the focus of the review.

#### ***4.4 Recommendations further research***

Findings from this realist review highlight the need for further research to support, refute or refine our proposed programme theory of how appraisal of doctors is supposed to produce its effects. Realist evaluation methods would be most suited to this endeavour. While a realist review depends on secondary data (i.e. published literature) a realist evaluation would gather primary data e.g. through interviews with appraisers and appraisees to 'test' the programme theory of how appraisal produces its effects. This type of research would generate findings to explain what occurs during appraisal meetings to instigate behaviour change and how contextual factors impact upon appraisal outcomes.

Once we fully understand how appraisal of doctors produces its effects the next step would be to understand how revalidation produces its effects. Again realist methods would be best to answer this research question. Revalidation is a separate intervention to appraisal although appraisal is part of the intervention strategy of revalidation. There was limited data in the literature we reviewed to support the idea of revalidation being an important context that impacts on appraisal. Although this was probably because there has been limited published research to date evaluating revalidation. Anecdotal evidence combined with ongoing empirical research currently being conducted will be able to fill this gap in the near future.<sup>21</sup>

## 5. CONCLUSION

This review is an important first step in understanding how appraisal of doctors produces its effects. It makes a significant contribution to the literature by identifying *dissonance*, *denial* and *self-affirmation* as three causal mechanisms at work in the appraisal process. The review also identified appraisers' skills, appraisees' characteristics, environment in which the appraisee works, and the relationship between the appraisee and appraiser as key contexts that are crucial to an effective appraisal.

## 6. REFERENCES

- [1] RST. Medical Appraisal Guide: A guide to medical appraisal for revalidation in England 2014. 16th April 2015. Available from:  
<http://www.revalidationsupport.nhs.uk/CubeCore/.uploads/documents/pdf/MAGpilotreport201112v1.1.pdf>.
- [2] Williams J. Appraisal for consultants. *Current Paediatrics*. 2003;**13**:549-554.

- [3] Mugweni K, Kibble S, Conlon M. Benefits of appraisal as perceived by general practitioners. *Education for Primary Care*. 2011;**22**:393-398.
- [4] General Medical Council. The state of medical education and practice in the UK report: 20152016 30th of June 2016. Available from: [http://www.gmc-uk.org/Chapter\\_1\\_SOMEPEP\\_2015.pdf](http://www.gmc-uk.org/Chapter_1_SOMEPEP_2015.pdf) [63501394.pdf](http://www.gmc-uk.org/Chapter_1_SOMEPEP_2015.pdf).
- [5] Shaw K, MacKillop L, Armitage M. Revalidation, appraisal and clinical governance. *Clinical Governance: An International Journal*. 2007;**12**:170-177.
- [6] Evans CC. Consultant appraisal. *Clinical Medicine*. 2003;**3**:495-496.
- [7] General Medical Council. Revalidation: A statement of intent2010 15th August 2015. Available from: [http://www.gmc-uk.org/Revalidation\\_A\\_Statement\\_of\\_Intent\\_October\\_2010\\_Final\\_version\\_web\\_version.pdf](http://www.gmc-uk.org/Revalidation_A_Statement_of_Intent_October_2010_Final_version_web_version.pdf) [35982397.pdf](http://www.gmc-uk.org/Revalidation_A_Statement_of_Intent_October_2010_Final_version_web_version.pdf) [36401645.pdf](http://www.gmc-uk.org/Revalidation_A_Statement_of_Intent_October_2010_Final_version_web_version.pdf).
- [8] Rubin P. Commentary: The role of appraisal and multisource feedback in the UK General Medical Council's new revalidation system. *Academic Medicine*. 2012;**87**:1654-1656
- [9] Health Do. Costs and Benefits: Analysis of costs and benefits of medical revalidation in England. Department of Health, London, 2012.
- [10] Archer J, Nunn S, De Bere SR, Anderson A. Revalidation in practice: shaping the future development of revalidation. Plymouth University2014 1st February 2015. Available from: [https://www.plymouth.ac.uk/uploads/production/document/path/1/1914/Revalidation\\_Revised\\_Stage\\_2\\_report\\_FINAL\\_09.07.14\\_1.pdf](https://www.plymouth.ac.uk/uploads/production/document/path/1/1914/Revalidation_Revised_Stage_2_report_FINAL_09.07.14_1.pdf).
- [11] Archer J. Revalidation built on appraisal may have led to confusion. *BMJ*. 2015;**351**.
- [12] Lewis M, Elwyn G, Wood F. Appraisal of family doctors: an evaluation study. *British Journal of General Practice*. 2003;**53**:454-460.

- [13] McKinstry B, Peacock H, Shaw J. GP experiences of partner and external peer appraisal: a qualitative study. *British Journal of General Practice*. 2005;**55**:539-543.
- [14] Chamberlain JM. Portfolio-based performance appraisal for doctors: A case of paperwork compliance. *Sociological research online*. 2010;**15**:8.
- [15] Scallan S, Locke R, Eksteen D, Caesar S. The benefits of appraisal: a critical (re) view of the literature. *Education for Primary Care*. 2016:1-4.
- [16] Merkur S, Mossialos E, Long M, McKee M. Physician revalidation in Europe. *Clinical medicine*. 2008;**8**:371-376.
- [17] MICHAEL G. TRAVELS OF A DOCTOR—A MEDICAL DRIVE THROUGH THE EU. *British Journal of General Practice*. 2009.
- [18] Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*. 2005;**10**.
- [19] Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. *BMC Medicine*. 2013;**11**:1-14.
- [20] Kastner M, Estey E, Perrier L, et al. Understanding the relationship between the perceived characteristics of clinical practice guidelines and their uptake: protocol for a realist review. *Implementation Science*. 2011;**6**:1-9.
- [21] UMBRELLA. UMBRELLA: shaping the future of medical revalidation2015 2nd of July 2016. Available from: <http://www.umbrella-revalidation.org.uk/>.
- [22] Archer J, Regan De Bere S, Brennan N, Lynn N. Evaluating the strategic impact of revalidation: Building an evaluation framework2013 1st of September 2015.

- [23] Archer J, de Bere SR, Nunn S, Clark J, Corrigan O. "No One Has Yet Properly Articulated What We Are Trying to Achieve": A Discourse Analysis of Interviews With Revalidation Policy Leaders in the United Kingdom. *Academic Medicine*. 2015;**90**:88-93.
- [24] Archer J, de Bere SR. The United Kingdom's experience with and future plans for revalidation. *Journal of Continuing Education in the Health Professions*. 2013;**33**:S48-S53.
- [25] Archer J, Regan De Bere S, Nunn S, Clarke J, Corrigan O. Revalidation: In policy 2012 2nd February 2016. Available from:  
<http://www1.plymouth.ac.uk/peninsula/research/camera/revalidation/Documents/Stage-One-Report.pdf>.
- [26] Walshe K, Boyd A, Bryce M, et al. Implementing medical revalidation in the United Kingdom: Findings about organisational changes and impacts from a survey of Responsible Officers. *Journal of the Royal Society of Medicine*. 2017:0141076816683556.
- [27] Brennan N, Bryce M, Pearson M, Wong G, Cooper C, Archer J. Understanding how appraisal of doctors produces its effects: a realist review protocol. *BMJ open*. 2014;**4**:1-5.
- [28] Ritzer G. *Metatheorizing in sociology* Lexington Books; 1991.
- [29] Roen K, Arai L, Roberts H, Popay J. Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Social science & medicine*. 2006;**63**:1060-1071.
- [30] Pearson M, Hunt H, Cooper C, Shepperd S, Pawson R, Anderson R. Intermediate care: a realist review and conceptual framework. *Final Report Southampton: NIHR Service Delivery and Organisation programme*. 2013.
- [31] Williams S, Pownceby P, Laugharne K. The impact of a yearly, structured appraisal on the general practitioner and their practice in Wales. *Education For Primary Care*. 2009;**20**.

- [32] McGivern G, Ferlie E. Playing tick-box games: Interrelating defences in professional appraisal. *Human Relations*. 2007;**60**:1361-1385.
- [33] West M. How can good performance among doctors be maintained?: Department of Health's proposals are wise but need to be implemented with care. *BMJ: British Medical Journal*. 2002;**325**:669.
- [34] Chamberlain JM. Portfolio-based appraisal: superficial or useful? *British Journal of Hospital Medicine*. 2009;**70**:614-615.
- [35] Nayar V. A qualitative study of general practitioners' views of the appraisal process. *Education for Primary Care*. 2005;**16**:672.
- [36] Jelley D, van Zwanenberg T. Practice-based peer appraisal in general practice: an idea whose time has come? *Education for Primary Care*. 2003;**14**:329-337.
- [37] Conlon M. Appraisal and revalidation: can the circle be squared? *Clinician in Management*. 2004;**12**:197-202.
- [38] Conlon M. Appraisal: the catalyst of personal development. *BMJ: British Medical Journal*. 2003;**327**:389.
- [39] Platt MW, Parkin J. Performance appraisal and career counselling for senior house officers. *Archives of disease in childhood*. 1990;**65**:796-801.
- [40] Eva KW, Regehr G. Effective feedback for maintenance of competence: from data delivery to trusting dialogues. *Canadian Medical Association Journal*. 2013;**185**:463-464.
- [41] Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;**6**:CD000259.

[42] Ivers NM, Sales A, Colquhoun H, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention.

*Implementation Science*. 2014;**9**:1-8.

[43] Brehaut JC, Eva KW. Building theories of knowledge translation interventions: Use the entire menu of constructs. *Implementation Science*. 2012;**7**:1-10.

[44] Brinko KT. The practice of giving feedback to improve teaching: What is effective? *The Journal of Higher Education*. 1993;**64**:574-593.

[45] Thurlings M, Vermeulen M, Bastiaens T, Stijnen S. Understanding feedback: A learning theory perspective. *Educational Research Review*. 2013;**9**:1-15.

[46] Neve H, Wearn A, Collett T. What are threshold concepts and how can they inform medical education? *Medical teacher*. 2015:1-4.

[47] Caesar S, Locke R, Scallan S. Medical appraisal and revalidation: can it improve patient care? *Education for Primary Care*. 2013;**24**:398-400.

[48] Griffin A, Furmedge DS, Gill D, et al. Quality and impact of appraisal for revalidation: the perceptions of London's responsible officers and their appraisers. *BMC medical education*. 2015;**15**:152.

[49] Adams R, Illing J, Jelley D, Walker C, van Zwanenberg T. The critical success factors in internal and external GP appraisal. *Education for Primary Care*. 2006;**17**:607-616.

[50] Zolle O, Rickenbach M, Gorrod E. The formative impact of general practice appraisals: a post-appraisal questionnaire survey. *Education for Primary Care*. 2009;**20**:445-454.

[51] Tavabie A, Koczwara A, Patterson F. Using emotional intelligence to facilitate strengthened appraiser development. *Education for Primary Care*. 2010;**21**:9-19.

- [52] Wakeling J, Cameron N, Lough M. What factors impact on general practitioners' engagement with appraisal? A qualitative evaluation in Scotland. *Education for Primary Care*. 2008;**19**:615-623.
- [53] Churchill D. Appraisal—a developing process. *Obstetrics, Gynaecology & Reproductive Medicine*. 2010;**20**:225-226.
- [54] Sapyta J, Riemer M, Bickman L. Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology*. 2005;**61**:145-153.
- [55] Chamberlain JM. *Doctoring medical governance: Medical self-regulation in transition*. New York Nova Science Publishers; 2009.
- [56] McGivern G, Adams R. Clinically owned reflective practice and tick box exercises: comparative analysis of GP and consultant appraisal in practice. *Clinician in Management*. 2006;**14**:129-141.
- [57] Hegde R, Hughson AM. Consultant appraisal of consultant appraisal. *Psychiatric Bulletin*. 2008;**32**:214-217.
- [58] Finlay K, McLaren S. Does appraisal enhance learning, improve practice and encourage continuing professional development? A survey of general practitioners' experiences of appraisal. *Quality in Primary Care*. 2009;**17**:387-395.
- [59] Conlon M, Sweeney G, Lyons N, Shelly M. Appraisal: experiences, attitudes and impact. An evaluation of the appraisal process for general practitioners in England. *Clinician in Management*. 2006;**14**:5-22.
- [60] Payne VL, Hysong SJ. Model depicting aspects of audit and feedback that impact physicians' acceptance of clinical performance feedback. *BMC Health Services Research*. 2016;**16**:260.

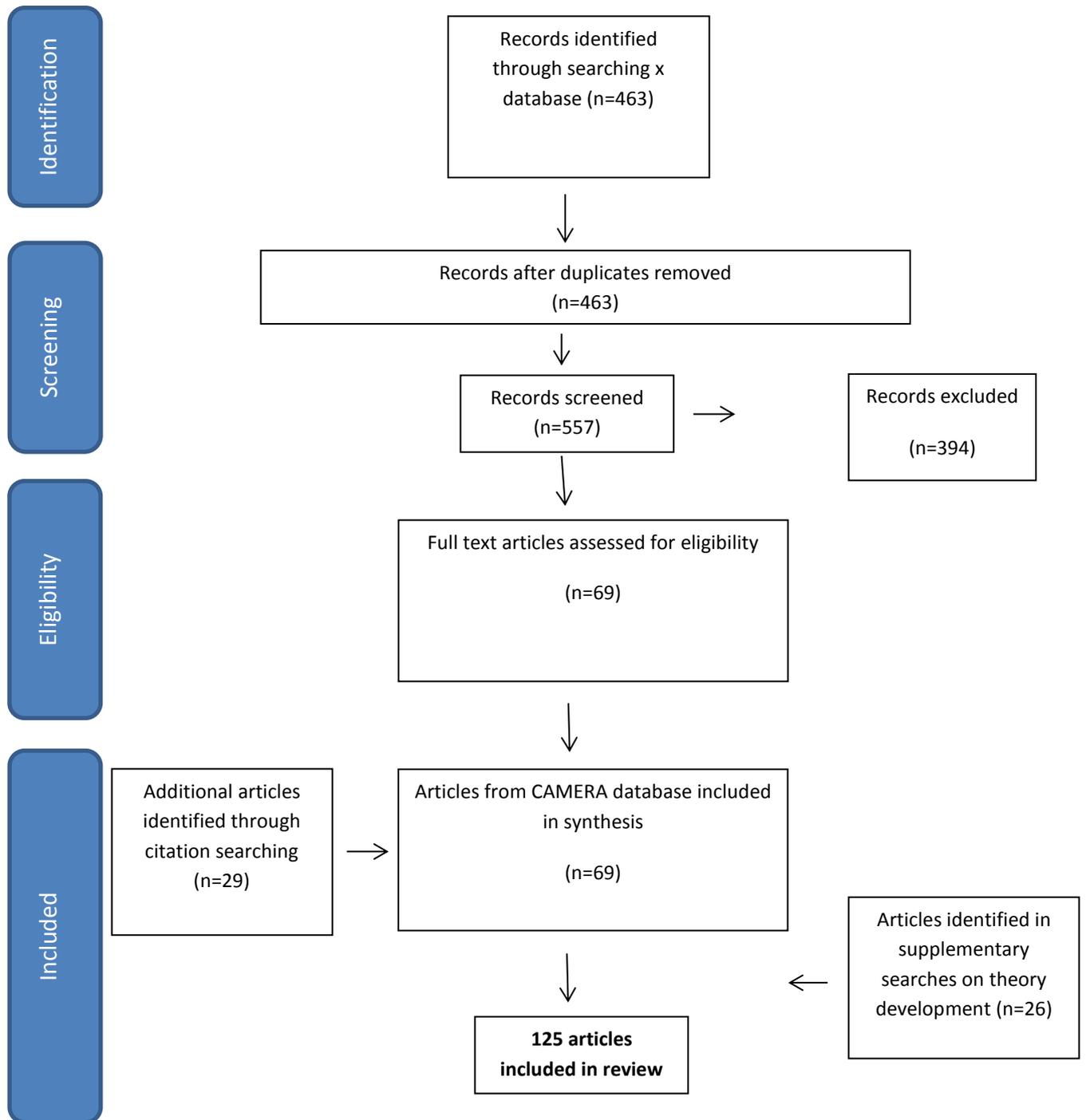
[61] Sherman DK. Self - affirmation: Understanding the effects. *Social and Personality Psychology Compass*. 2013;**7**:834-845.

[62] Cohen GL, Sherman DK. The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*. 2014;**65**:333-371.

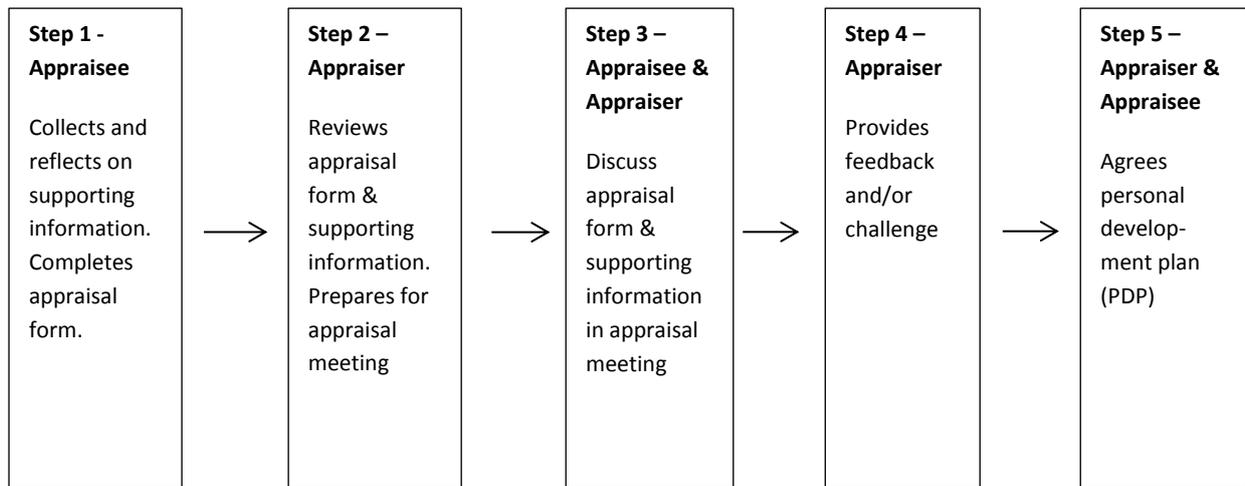
[63] Vince R. Behind and Beyond Kolb's Learning Cycle. *Journal of Management Education*. 1998;**22**:304-319.

[64] Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychol*. 2008;**27**.

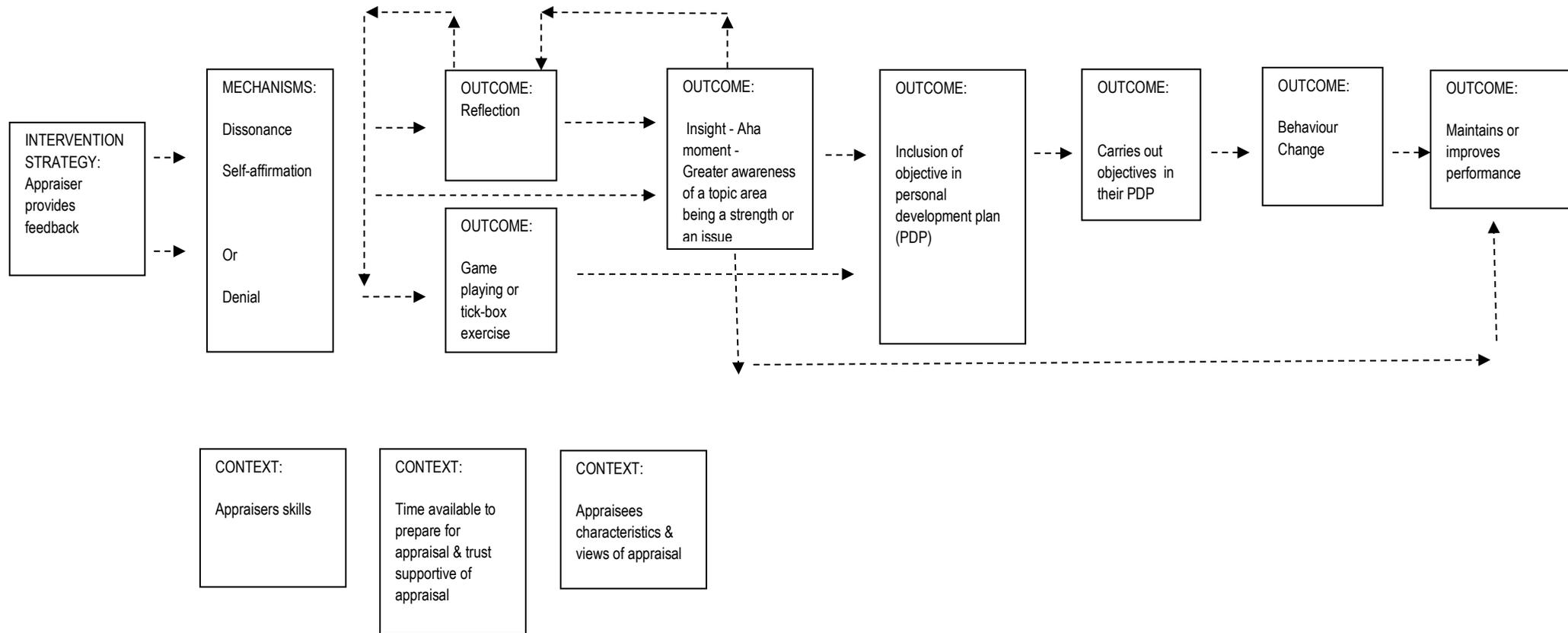
**Figure 1 Flow Chart of Screening Process**



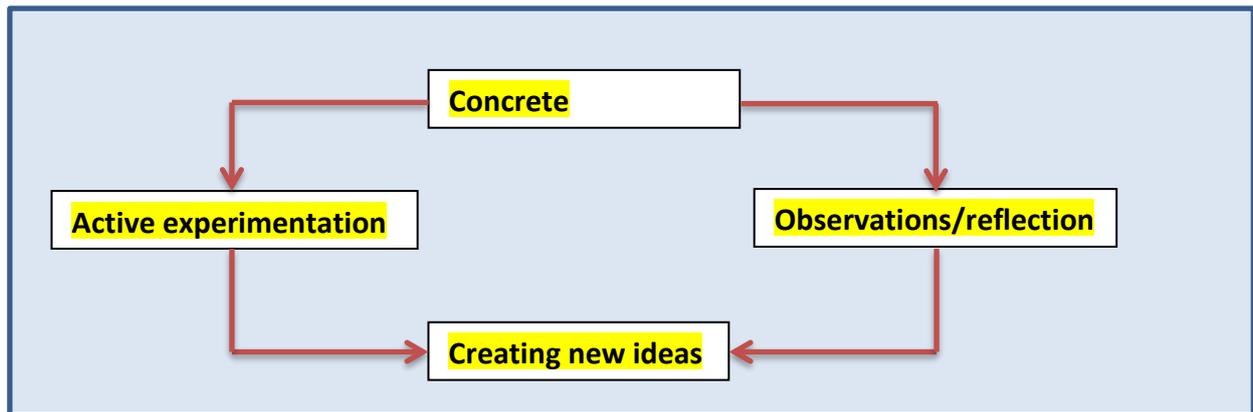
**Figure 2: Steps of Appraisal Intervention Strategy**



**Figure 3 Programme Theory of Appraisal of Doctors**



**Figure 4: Learning Cycle (in Conlon adapted from Kolb)**



**Table 1 Criteria used for assessing the conceptual richness of sources**

'Conceptually rich'	'Thicker description' but not 'conceptually rich'	'Thinner description'
Theoretical concepts are unambiguous and described in sufficient depth to be useful	Description of the programme theory or sufficient information to enable it to be 'surfaced'	Insufficient information to enable the programme theory to be 'surfaced'
Relationships between and amongst concepts are clearly articulated	Consideration of the context in which the programme took place	Limited or no consideration of the context in which the programme took place
Concepts sufficiently developed and defined to enable understanding without the reader needing to have first-hand experience of an area of practice	Discussion of the differences between programme theory (the design and orientation of a programme—what was intended) and implementation (what 'happened in real life')	Limited or no discussion of the differences between programme theory (the design and orientation of a programme-what was intended) and implementation (what 'happened in real life')
Concepts grounded strongly in a cited body of Literature	Recognition and discussion of the strengths and weaknesses of the programme as implemented	Limited or no discussion of the strengths and weaknesses of the programme as implemented
Concepts are parsimonious (i.e. provide the simplest, but not over-simplified, explanation)	Some attempt to explain anomalous results and findings with reference to context and data	No attempt to explain anomalous results and findings with reference to context and data
	Description of the factors affecting implementation	Limited or no description of the factors affecting implementation
	Typified by Terms—'model', 'process', or 'function' Verbs—'investigate', 'describes', or 'explains' Topics—'experiences'	- Typified by Mentioning only an 'association' between variables

**Table 2 Characteristics of core literature of appraisal of doctors**

<b>Study Characteristics</b>	<b>Number and proportion of studies</b>
<b>Article source</b>	
Education for Primary Care	n=38/96 (39.5%)
British Journal of General Practice	n=11/96 (11.4%)
Clinician in Management	n=9/96 (9.4%)
Report	n=8/96 (8.3%)
≤3 articles in 25 other journals	n=30/96 (31.2%)
<b>Country article related to</b>	
UK	n=77/96 (80.2%)
Scotland	n=11/96 (11.4%)
Wales	n=4/96 (4.1%)
Northern Ireland	n=3/96 (3.1%)
<b>Healthcare setting</b>	
Primary care	n=57/96 (59.3%)
Secondary care	n=13/96 (13.5%)
Primary and secondary care	n=26/96 (27.0%)
<b>Focus of paper or article</b>	
Summary of appraisal process	n=18/96 (18.7%)
Perceptions of appraisal	n=13/96 (13.5%)
Appraiser training	n=11/96 (11.4%)
Impact of revalidation on appraisal	n=6/96 (6.3%)
Impact of appraisal	n=6/96 (6.3%)
Appraisal implementation	n=4/96 (4.1%)
≤3 articles on 24 other topics of appraisal	n=42/96 (43.8%)
<b>Methods utilised</b>	
Opinion piece/editorial	n=35/96 (36.4%)
Qualitative	n=19/96 (19.7%)
Questionnaire survey	n=15/96 (15.6%)
Description of appraisal process	n=7/96 (7.2%)
Mixed methods	n=14/96 (15.7%)
Literature review	n=5/96 (5.2%)
<b>Quality assessment</b>	
Thinner description	n=78/96 (81.2%)
Thicker description	n=15/96 (15.6%)
Conceptually rich	n=3/96 (3.1%)