

Using narrative evidence synthesis in HRM research: an overview of the method, its application and the lessons learned.

Abstract

The use of systematic approaches to evidence review and synthesis has recently become more common in the field of organizational research, yet their value remains unclear and largely untested. First used in medical research, evidence review is a technique for identifying, evaluating and synthesizing existing empirical evidence. With greater demand for the best evidence about ‘what works’ in organizational settings, nuanced approaches to evidence synthesis have evolved to address more complex research questions. Narrative synthesis is perceived to be particularly suited to evaluating diverse evidence types spanning multiple disciplinary fields, characteristic of the HRM domain. This article evaluates the narrative evidence synthesis approach, explains how it differs from other techniques and describes a worked example in relation to employee engagement. We consider its strengths, the challenges of using it and its value in HRM research.

Introduction

Doing organizational research is not easy: determining what counts as reliable evidence in multi-disciplinary areas of organizational research such as human resource management (HRM) has become increasingly difficult. It is a field characterized by complex evidence that includes data derived from the use of multiple and varied methods, often concerning new, emerging and contested topics. This conceptual and methodological complexity creates problems when trying to provide evidence and insight on best practice. It also generates uncertainty for researchers who are trying to understand the weight and trajectory of existing research in order to inform thinking about research gaps and future research directions. For example, Rousseau, Manning & Denyer

(2008: 476) comment on the ‘failure of Management and Organizational Science to date to make full effective use of its available research evidence. Failure to make *effective use* of scientific evidence is a problem both management scholars and practitioners face.’

It has been argued that one reason for the lack of clear and reliable evidence is that the traditional method of literature review, widely used in organizational research, frequently fails to grasp the nature and scope of complex evidence, or to provide the clarity and insight needed to aid conceptual understanding or practical decision-making (Rousseau, 2006). Although the traditional literature review remains a cornerstone of scholarly research, there is a concern that it often reduces, oversimplifies and misrepresents complex topics such as those found in HRM, and unduly excludes crucial or nuanced evidence that might be of interest and relevance (Morrell & Learmonth, 2015; Rousseau et al., 2008). Systematic approaches to evidence synthesis, on the other hand, have evolved since they first emerged in the medical field to offer more rigorous approaches for assessing and synthesizing complex evidence (Greenhalgh, Glenn, Bate, Kyriakidou, MacFarlane, et al., 2004), and thus offer an alternative, and more structured approach to capturing and evaluating the entire evidence base relating to a specific topic.

Despite the availability of these more systematic approaches to evidence synthesis, and a growing movement that advocates their wider use, Patterson, Rick, Wood, Carroll, Balain, et al., (2010: 7) observe that they still have a very ‘limited pedigree within HRM’. In this article, we consider the main approaches to evidence review that have evolved in the field of organizational research, setting out the arguments for their use and highlighting their perceived strengths and limitations. We focus in particular on one such structured approach to evidence review, *narrative evidence synthesis*, and describe and evaluate a worked example of this approach on the important and contested topic of employee engagement to illustrate the practical, logistical and methodological

challenges it posed. Finally, we explain its benefits and make recommendations for future research in the HRM field.

Emergence of evidence review

The idea of evidence review is not new. The move towards using evidence reviews first emerged in the field of medical research in the 1970s, arising from a perceived need to devise techniques for the systematic review and synthesis of evidence from different studies to ensure that decisions about patient care and treatment were based on best available evidence (Denyer & Tranfield, 2011; Boaz, Ashby, Denyer, Egan, Harden et al., 2006). It quickly spread into health care management in the 1980s through a growing interest in ‘evidence-based policy’ and ‘evidence-based practice’, and the desire to access the best information about ‘what works’ in order to inform decision-making about resource allocations (Morrell, 2008; Tranfield, Denyer, & Smart, 2003). These emerging approaches all sought to integrate expertise (‘proficiency and judgment’) with the ‘best available’ (relevant and reliable) evidence drawn from the systematic review of studies within a given field in order to minimize the effect of bias from individual or small numbers of studies (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996: 71).

By the 1990s, the movement became formalized in the UK through the development of clear guidelines for the conduct of systematic reviews, firstly in the field of medicine through the Cochrane Collaboration and, later, in the social sciences through the Campbell Collaboration (Briner & Denyer, 2010). Its goal was to promote and coordinate the availability of reliable evidence to achieve greater transparency and legitimacy of decisions based on good evidence and rational judgments (Morrell & Learmonth, 2015). Over time, as this movement has expanded to include wider policy and organizational settings in social care, education and management, it has

created the expectation that all research and actions should be evidence-based (Denyer & Tranfield, 2011). Outside the field of medicine, however, there has not been a clear consensus on what ‘relevant’ and ‘reliable’ evidence actually means, or how to evaluate it (Viswanath & HakemZadeh, 2012).

Need for evidence-based reviews in organizational research

In management and other disciplines in the social sciences, the volume of information and research has increased exponentially over the past 20 years (Badger, Nursten & Woodward, 2000). Unlike research in the medical field, topics in fields such as HRM are characterized by a range of information and data that reflect different organizational realities and where the evidence is much less homogeneous (Pawson, Greenhalgh, Harvey & Walshe, 2004). Information is presented not simply in terms of data from experimental research but also in the form of complex statistical analyses, rich case studies, biographies and ethnographic research about human agency set within their wider social and cultural contexts. The evidence is thus heterogeneous. As Buchanan & Bryman (2011: xxv) state, the diverse and ‘eclectic’ nature of this discipline has not only led to new areas of research, research methods, and a new lexicon of research, it also includes ... ‘the collection of non-traditional forms of evidence, the development of fresh approaches to measurement, conceptualization and theory building, and the acceptance of modes of generalizability of findings other than statistical’. Within this wide range of research approaches, data can of course be of variable and sometimes dubious quality, even when published in high quality journals and, at the extreme, feature inappropriate study designs and analysis.

This heterogeneity makes rational judgments about what constitutes reliable evidence and how to assess it somewhat more challenging in the organizational field than in medicine. Evidence about organizational management often reflects the ‘politics of organizational action’ and ‘the interaction between managerial agency/choice and structural constraints and opportunities in the shaping of human resourcing practices’ (Watson, 2004: 450). Within established topics of HRM research such as employee engagement, as well as new or emerging topics such as employer branding or ‘green’ HRM, the evidence commonly spans multi-disciplinary interests including psychology, sociology, health, organization studies, human resource management and management studies, drawing on a variety of data types, methods and analytic frameworks (App, Merk, & Büttgen, 2012). Consequently, determining what constitutes best evidence in the HRM field is problematic, not only because conceptual complexity makes it more difficult to ask the kind of clearly specified questions typical in medical research, but also because it makes the development of appropriate technologies to enable the robust examination, analysis and coherent summary of research on such topics much harder to achieve. In light of these issues, it has been argued that the traditional approach to literature review cannot provide the rigor or the reliability necessary to assess and evaluate the evidence base. As Rousseau et al. (2008: 475) state, ‘traditional literature reviews are often position papers, cherry-picking studies to advocate a point of view’, and do not offer researchers a framework for undertaking a thorough and systematic evaluation of the evidence (Briner & Denyer, 2010). We consider this argument further in the next section.

Traditional approaches to review

The traditional literature review approach is an examination of knowledge on a given topic using appropriate sources of information (Dawidowicz, 2010). This examination usually leads to a

summary of the existing state of knowledge, identifying gaps or contradictions in the underpinning theory and data from which key questions and hypotheses for further research can be articulated (Jones & Gattrell, 2014; Greenhalgh et al, 2004). Its ubiquity has ensured that it remains the cornerstone of scholarly research, yet it suffers from a number of limitations, especially when the focus is on reviewing the evidence relating to topics that span multiple disciplinary interests with complex evidence bases (Denyer & Tranfield, 2005; Popay, Roberts, Sowden, Pettigrew, Arai et al., 2005).

Although widespread in use, there is no prescribed 'systematic' approach to traditional literature reviews. They vary from a simple, rapid summary on a given topic to a more applied, in-depth review of complex issues and although this makes the traditional literature review a very flexible tool, the approach – including structure and style - varies according to scholarly discipline as well as the individual, expert preferences of those who undertake or guide them, along with institutional expectations (Briner & Denyer, 2010). What gets included in the resulting literature review may therefore be susceptible to problems of researcher and publication bias (Hammersley, 2001). These biases may reflect tacit disciplinary preferences for certain types of data or methods, expert preferences for underpinning theories, argument structure and techniques of data analysis, or the editorial preferences of academic journals and conferences (Ecklund & Scheitle, 2007). Furthermore, as Rousseau et al. (2008: 477) argue, 'the explosion of management research since World War II has created knowledge products at a rate far outpacing our current capacity for recall, sense-making, and use', casting doubt on the ability of the traditional literature review approach to thoroughly and accurately account for the volume and heterogeneity of evidence on a given topic.

Traditional literature reviews may commence with a clear question to be addressed and develop through general discussion of a topic, but they rarely evaluate why certain parts of literature are included or excluded, or explore why the general direction of research on the focal topic may have changed, potentially leading to misinterpretation or misrepresentation of the evidence (Dixon-Woods, Cavers, Agarwal, Annandale, Arthur, et al., 2006; Pettigrew, 2001).

Consequently, the summaries produced through the traditional literature review approach are deemed to be inexhaustive and prone to subjectivity (Slavin, 1996), and ‘highly problematic when the evidence upon which they are based is subjected to closer scrutiny’ (Hodgkinson & Ford, 2014: S2). Viswanath & HakemZadeh (2012: 832) claim that those trying to make sense of the summaries produced by these reviews, including practitioners and policy-makers, are often left to do so on the basis of ‘gut feelings’ and instinct rather than on the clarity or comprehensiveness of the evidence.

Systematic approaches to evidence synthesis

In response to these limitations, there has been a call for more systematic approaches to analyzing the literature (Slavin, 1995). Such approaches provide what Tranfield et al. (2003: 209) describe as ‘a detailed technology that aims to minimize bias through exhaustive literature searches ... by providing an audit trail of the reviewers’ decisions, procedures and conclusions’. These systematic approaches have enabled the assembly, analysis and synthesis of disparate research data in order to clarify the breadth and depth of the evidence base on the focal topic. Moreover, the insights arising from systematic review enable the generation of new research questions or hypotheses that would not otherwise have been possible (Humphrey, Nahrgang & Morgeson, 2007; Boaz et al., 2006; Britten, Campbell, Pope, Donovan, Morgan et al., 2002). It has been argued that the methodological soundness and emphasis on quality in such reviews are more

powerful than those resulting from traditional literature reviews since their approach has led to greater validity and confidence in the results (Briner & Denyer, 2010).

Is evidence review bias free?

Although systematic reviews have been welcomed as offering ‘the best way forward’ in evaluating complex evidence from diffuse sources (Rousseau et al., 2008: 8), others question whether the claims of minimizing bias and subjectivity are in practice so well-founded. Lawler (2007) and Learmonth & Harding (2006) suggest that all approaches to evidence review denote specific ‘organizing’ (e.g. ‘medical’ or ‘management’) narratives that, by definition, embody a particular way of valuing evidence. Hence, the whole evidence-based movement, it is argued, shares a particular (‘epistemic’) worldview based on scientific and rationalist values that is anything but ‘value-free’ (Morrell, 2006: 616). Thus, the evidence-based approach maybe ‘itself a story about relations between research and practice, one of many possible stories’ (Morrell & Learmonth, 2015: 521). Topics such as those in HRM abound with such relational complexity that the desire to use evidence-based techniques appear to some little more than ‘wistful’ (Morrell & Learmonth, 2015: 620). The diffuse and voluminous nature of the research evidence in HRM points to the existence of multiple and competing bodies of evidence whose breadth of field and underlying values may defy summation or generalization, rendering the task of systematic analysis and synthesis problematic (Learmonth & Harding, 2006). Despite these concerns, the perceived success of evidence-based approaches within the field of medicine has heightened their appeal in the social sciences (Rousseau, 2006). The movement advocating the use of evidence review techniques has therefore spread, becoming ever more popular in management and organizational research, even though there is to date a paucity of evidence that they necessarily lead to better outcomes or insights (Viswanath & HakemZadeh, 2012).

As interest in evidence reviews has evolved from the narrow context of systematic reviews in medicine first advocated by Cochrane (1972), their adoption into new disciplines has prompted the development of a range of different and more nuanced approaches that reflect the divergent nature of knowledge in these disciplines (Boell & Cecez-Kecmanovic, 2011; Tranfield et al., 2003). Following the underlying principle of procedural rigor established by Cochrane (1972), Gough, Oliver & Thomas (2013) state that even if alternative approaches to evidence review have emerged beyond medicine, they must still adhere to the same basic logic, rules and conduct of systematic review. Before considering the relevance, value and practicality of undertaking an evidence synthesis, we describe the five stages of the systematic review process in the next section.

What does an evidence review involve?

Any review of the evidence, whether traditional or systematic, should be both thorough and explicit in order to demonstrably reduce bias (Cochrane, 1972; Sackett et al., 1996). However, since the methodological procedures are explicit, the conduct of an evidence review is replicable and has become codified through five distinct stages (Briner, 2011):

- (i) *planning the research strategy and developing the research questions*
- (ii) *searching for and locating evidence*
- (iii) *sifting search results for studies that meet the agreed criteria*
- (iv) *systematically extracting data from included studies and evaluating them, and*
- (v) *synthesizing conclusions, making recommendations and disseminating findings.*

Based on these five steps, evidence review has thus become defined as a standard methodology that identifies and locates empirical studies relevant to an agreed research issue, evaluates these

studies and then synthesizes their data, reporting on what is found ‘in such a way that allows reasonably clear conclusions to be reached about what is and is not known’ (Denyer & Tranfield, 2011: 671).

There are three notable aspects of this process. Firstly, the initial planning stage seeks to ensure that evidence reviews address ‘precise, answerable and meaningful’ research questions as well as devising an effective strategy for finding the evidence (Briner & Denyer, 2010: 344). Popay et al. (2006: 9) state: ‘Getting the question(s) “right” is critical to the success of the systematic review process overall’. To facilitate this, the Cochrane Collaboration (2010) recommend that review questions, as well as the conduct of the review, be tested, developed and guided with the help of an advisory group, an approach that first emerged in the medical field to try to ensure that the reviews were conducted in a judicious and transparent manner. Comprising a panel of topic experts, practitioners or other representatives (e.g. service users, policy makers) relevant to the topic in hand, it is argued that advisory groups can add particular value to evidence reviews in social science research. Higgins & Green (2008) add that review findings are likely to be of higher quality and to have greater relevance if the parameters and conduct of the review process have been guided by people who have expert, practical or methodological insight on the topic. In addition to ensuring transparency, the involvement of experts and practitioners in supporting research teams to assess complex HRM topics can help to ensure that reviews are of sufficient scope and clarity to be meaningful.

Secondly, while the logic of reviews should always follow these five steps, the conduct of reviews in management and other research has progressively evolved to avoid over-adherence to strict rules that might constrain opportunities to respond creatively and meaningfully to the evidence itself, given its complex and problematic nature (Briner & Denyer, 2010). Thirdly,

while the overall methodology is standardized, the specific approach to the review of the evidence varies according to the type of synthesis used. In the next section we consider the different approaches to evidence review and highlight some of their perceived strengths and limitations.

Approaches to evidence review and synthesis

Rousseau et al. (2008) suggest that while different approaches to evidence synthesis have emerged, they generally fall into one of four types: *aggregative*, *integrative*, *interpretative* and *explanatory*, each having particular strengths in considering certain types of research topic and evidence. To these, a fifth distinctive approach – *narrative evidence synthesis* – has been added which is seen to offer the greater sensitivity and reflexivity necessary for reviewing complex topics and data types such as those found in HRM research (Greenhalgh et al., 2004). We outline these five different approaches in Table 1, starting with the *aggregative* approach which is perhaps most commonly associated with the traditional systematic reviews that first emerged in the medical field, and from which the other approaches have evolved. Based on the five steps outlined above, the ‘detailed technology’ of systematic review (Tranfield et al., 2003: 209) involves an exhaustive search of literature in relation to a specific and clearly articulated question. Once studies are identified and their primary data extracted, the results are aggregated in order to provide a reliable response to the review question (Denyer & Tranfield, 2011; Tranfield et al., 2003).

INSERT TABLE 1 ABOUT HERE

Outside the field of medical research, other more sophisticated and powerful statistical techniques have emerged to assist in the process of combining and analyzing large data sets, most commonly

in the form of meta-analyses (Humphrey et al., 2007). These are seen to be useful where, in addition to meeting criteria of quality and relevance, studies included in meta-analyses involve statistical data on the same phenomenon that can be aggregated, compared and synthesized to identify the effectiveness of an intervention or the strength of relationship between variables (Tranfield et al., 2003). It is an approach that has some limitations however, particularly when seeking to review evidence on complex research questions or to aggregate complex datasets. The demand for standardized data as the basis of aggregation and analysis means these reviews are often based on fewer studies than might appear to be the case (Denyer & Tranfield, 2011).

Integrative synthesis provides a slightly different approach to evidence review in that it includes quantitative and qualitative data from different studies, but the included data must still meet equivalence criteria for quality (Briner & Denyer, 2010). Emerging in the field of healthcare and nursing management, integrative reviews are based on the recognition that evidence concerning the effectiveness of medical interventions needs to include data on the experience of service users, families and others, often excluded from aggregative reviews. As Whitemore & Knafl (2005: 547) state, by including evidence from multiple sources and different data types, integrative reviews widen the scope of the systematic approach to include ‘diverse methodologies ... and contribute to the presentation of varied perspectives on a phenomenon of concern’. Since they have wider and more complex evidence bases, integrative reviews enable multiple rather than single review questions to be asked, and are appropriate for exploring new and emerging topics about which there is a limited amount of data or where there is little convergence on acceptable research methods and conceptual underpinnings. There are many instances when integrative syntheses have produced ‘seminal contributions to knowledge’ (Torraco, 2005: 356) because, in addition to reviewing and synthesizing empirical data, they enable evidence to be

further maximized by including assessment of the theoretical frameworks on which studies are based (Marler & Fisher, 2013). A perceived problem of the integrative approach, however, is that it is not as straightforward to integrate data or theoretical frameworks emanating from both quantitative and qualitative studies as it might appear, since they are generated from different assumptions about the constitution of knowledge (Briner & Denyer, 2010). This is exacerbated because there is rarely a comparable volume of quantitative and qualitative data available in organizational research, where the weight of evidence is commonly imbalanced towards the quantitative.

The third approach to evidence review is the *interpretive* method of evidence review and synthesis. With its roots within the hermeneutic tradition of social research, interpretive reviews move through the five steps as described above but develop the emerging synthesis through induction and interpretation, with the goal of generating or critiquing theory rather than testing it. (Dixon-Woods et al., 2006). The approach is seen to be useful when exploring established concepts and issues with large but complex evidence bases and where more critical assessment is needed (Barnett-Page and Thomas, 2009). It has particular value in asking questions beyond the effectiveness of interventions to consider issues of ‘intervention need, appropriateness and acceptability, and factors influencing intervention implementation’ (Thomas & Harden, 2008: 2). Drawing on different techniques often used in qualitative research, such as reciprocal translational analysis, interpretive reviews aim to synthesize study findings and their conceptual underpinnings by ‘translating’ their meaning into higher order analytical themes, systematically comparing emerging findings from each study, leading to the reconceptualization of a topic based on the evidence (Thomas & Harden, 2008; Dixon-Woods et al., 2006).

It is often claimed that interpretive reviews are concerned solely with qualitative research (Campbell, Pound, Morgan, Daker-White, Britten et al., 2011), but Dixon-Woods et al. (2006) state that a well-conducted interpretive review can include any evidence, regardless of study type. However, the concern is less about delimiting strict criteria for including studies because this often means good studies with important implications are excluded on the basis of ‘surface mistakes’ (Dixon-Woods et al., 2006). As a result, interpretive reviews draw similar criticisms of subjectivity and bias compared with those leveled at traditional literature reviews (Briner & Denyer, 2010), and for lacking an agreed, standardized methodology (Dixon-Woods et al., 2006; Britten et al., 2002). Their emphasis on critical interpretation and reconceptualization would moreover suggest they are academically-oriented and less appealing to policy and decision-makers whose evidence needs and interests may be more instrumental.

The fourth approach - *explanatory evidence review* - is described as a ‘realist’ approach in that the approach of identifying causal (conjunctive) relationships between variables by aggregating and analyzing study data is rejected. Instead, explanatory reviews are perceived as more practical in nature, considering the underlying theories and assumptions of studies, the explanations and interpretations of researchers undertaking them, as well as the context in which those interpretations arise. Pawson et al., (2004: iii) state that explanatory reviews have value because they are more ‘compatible with the complexities of modern health service delivery and sympathetic to the usage of a multi-method, multi-disciplinary evidence base’. By focusing on the underlying theories of interventions, the aim of explanatory reviews is to produce a nuanced and pragmatic synthesis based on the ‘complex pattern of success and failure found in the empirical evidence’, rather than summary, critique or reconceptualization (Pawson, 2004: 1). They are as concerned with the question of ‘what doesn’t work?’ as with what does work, and

with mapping the inter-connected, complex contexts and circumstances in which interventions are successful or otherwise. In this sense, they are ‘realist’ because they approach interventions ‘as complex [social] systems thrust amidst complex systems’ (Pawson et al., 2004: iv). More conventional approaches to systematic review are not configured to consider or respond to the complexity of evidence or indeed the complexity of insight that explanatory review seeks. Pawson et al. (2004: iv) however state that one of the drawbacks of this approach to evidence review is that the process of ‘scouring’ for and synthesizing evidence is a ‘hard slog’ while the resulting explanation can be as complex as the topic under analysis. This makes it less appealing to decision-makers or practitioners than conventional reviews which aim to provide consensus and ‘homogenised’ answers to research questions.

The fifth approach to evidence review which aims to overcome many of the weaknesses of other approaches without compromising on precision, but which utilizes the flexibility of the traditional narrative literature review, is *narrative evidence synthesis*. This is seen to be of particular relevance to HRM research because of the capacity of narrative synthesis to bridge the divides between research, practice and policy (Briner & Denyer, 2010; Popay et al., 2006). Notably, this is true in relation to topics that have reached a level of maturity, with a considerable volume of available evidence but a lack of coherence or consensus about the topic and lack of clarity about causality. Narrative synthesis is based on the five steps of systematic review but differs in a number of ways. Firstly, fundamental to the *narrative* approach is the idea of being as inclusive and flexible as possible from the outset, including permitting the organizing narrative of the evidence to develop during the conduct of the review, rather than being fully established from the outset (Briner & Denyer, 2010). This means taking extra time and care at the beginning to consider all possible disciplines, sources and types of evidence, including the grey literature (i.e.

outputs that have not necessarily been subject to peer-review)¹ in order to identify relevant and underlying concepts relating to the topic under review. Should ideas later emerge about scope and content, these must be absorbed into the review as transparently and reflexively as possible to form part of the overall narrative.

Secondly, this approach questions whether the assumed values of rigor and validity are the only bases on which to determine whether evidence is of sufficient quality to be included, especially when dealing with complex topics involving people and organizations (Kitchenham, 2004).

Instead of seeking aggregation or integration, narrative synthesis means listening and attending to the conflictual nature of the evidence as a distinctive and important feature of the narrative, rather than as a conceptual weakness or empirical anomaly to be excluded.

Thirdly and following from this, allowing different types of data to be used helps to develop a descriptive yet critical narrative about a given topic, offering a method for embracing a wide body of disparate evidence with the aim of “telling the story” of the findings (Popay et al., 2006: 1). Being narrative in orientation means exploring how words and text as well as data are used in a range of diverse contexts in order to develop an integrated narrative overview of the evidence base and how this has unfolded over time (Kitson, Athlin, Elliott, & Cant, 2014; Briner & Denyer, 2010; Popay et al., 2006). This temporal aspect of narrative synthesis helps to distinguish it from other forms of evidence review.

Finally, while narrative synthesis seeks to tell the story of the evidence, it does not seek homogeneity or resolution of explanation in the form of a *happy ending*; given the complexity and maturity of the topic, the approach expects that the evidence may contain multiple strands; what it seeks is to produce a plausible account of this plurality rather than an answer or an

explanation as is sought by other approaches. As Greenhalgh, Robert, Macfarlane, Bate, Kyriakidou et al. (2005: 427) note: ‘If the body of evidence is complex, there will be no simple, formulaic or universal “solution” hidden in the literature awaiting discovery, nor will a single theory explain all findings ... The challenge is to expose the tensions, map the diversity and communicate the complexity...’. This makes narrative synthesis an important but potentially problematic methodology because it signifies an explicit return to the narrative approach and the concern that, no matter how systematic the underpinning process of review, the risk of bias and subjectivity still remains.

Of the five approaches to evidence review described above, the narrative synthesis approach is perhaps the least well established or utilized. Consequently, neither the claims of rigor nor the criticisms of subjectivity have been subject to extensive assessment, despite their continued expansion into the social sciences (Boell & Cecez-Kecmanovic, 2011). In the next section, we consider specifically how the narrative synthesis approach to evidence contributes to knowledge through a worked example on the topic of *employee engagement*.

Narrative synthesis of employee engagement

Like many topics in HRM research, employee engagement is a complex and contentious concept (Wefald & Downey, 2009; Purcell, 2014). In his seminal article, Kahn (1990: 700) defined personal role engagement as ‘the simultaneous employment and expression of a person’s “preferred self” in task behaviors’. Kahn’s interests lay primarily in people’s work motivation expressed along physical, cognitive and emotional dimensions but, since then, others have developed the concept in the directions of work attitudes, personality traits, emotional states and behavior, as well as organizational commitment, advocacy and involvement (Alfes, Shantz, Truss

& Soane, 2013; Guest, 2014). Particularly from 2003 onwards, there has been a rapidly expanding and diverse body of literature on engagement from both scholarly and practitioner perspectives, leading to a wide range of different definitions and measures (Fletcher et al., 2014; Crawford et al., 2014). According to Wilkinson and Redman (2013: 9), engagement has taken root in the HRM field because it has been shown that higher levels of engagement are positively associated with employee morale, wellbeing and organizational performance (Hu & Schaufeli, 2011; Rich, Lepine & Crawford, 2010). However, its prevalence in management and organizational research as well as in practitioner discourse has resulted in it being likened to a 'fad', about which much has been published yet about which there is seemingly little conceptual clarity (Guest, 2014). The burgeoning literature on engagement shows that there is considerable disagreement about the construct's purpose, definition and dimensionality (Schaufeli & Salanova, 2011; Truss, Alfes, Delbridge, Shantz & Soane, 2014), suggesting it is a topic that has reached a point where narrative synthesis could make a strong contribution to the development of the field.

Prior to our study, there had been no synthesis that brought together the wider evidence base relating to engagement. Our investigation was produced in response to a call for a narrative synthesis of the evidence on engagement by a national research agency in the health care field that had become interested in claims about the capacity for employee engagement to address problems in the health organizations around low morale, staff performance and work dissatisfaction. Although there were some summary literature reviews available (e.g. Shuck, 2011) many were procedurally unclear, and did not provide a critical account of the weighting of evidence discussed. The national health agency commissioning the evidence review was concerned that advice given to managers and other practitioners may be based on research that

demonstrated persuasive yet spurious correlations and linkages, rather than on rigorous, academic research grounded in theory and high-quality data.

We describe how we set about answering this call, but our goal is not to explain in depth what each stage of the review entailed; there are many articles which provide this technical detail concerning the conduct of an evidence review (for example, see Briner & Denyer, 2010; Popay et al., 2006). Rather, we aim to highlight the practical issues that arose in the process of carrying out the narrative synthesis and to show how, despite raising many logistical problems, it led to the development of insights that might not otherwise have been possible had we used the traditional literature review approach.

Narrative synthesis: using the method

Based on the guidelines set out by Briner (2011) we followed the five steps of evidence review, adhering to the key principles of organization, transparency, replicability, quality, credibility and relevance, as follows:

Initial planning

In line with the guidance on Cochrane reviews (Higgins & Green, 2008), we recruited a large inter-disciplinary research team and an Advisory Group comprising topic and methodological experts, practitioners, representatives from unions, employers and others. The group actively helped to guide the conduct of the review throughout, including scoping out the parameters of the review, identifying sources of evidence and the search strategy, contributing to the development of the research questions and advising on emerging findings and dissemination. The research questions that were developed to guide the narrative synthesis focused on definitions, theories, antecedents and outcomes of engagement and are listed in Table 2.

INSERT TABLE 2 ABOUT HERE

After developing a detailed search strategy, a preliminary scan of the literature on engagement identified 712,550 items from diverse sources. Using a CIMO framework as recommended by Denyer and Tranfield (2009) to develop research questions and the search strategy, we analyzed these items and their sources with reference to the *Context* (such as discipline or publication type, purpose, etc.) in which items had been generated; *Interventions* that had been evaluated in terms of expected causal relationships; *Mechanisms* through which these relationships were anticipated to lead to outcomes (including moderators and mediators), as well the types of *Outcomes* described (such as performance or morale). Based on this, we devised a search string of key terms in consultation with the advisory group which was then refined into a short search string that was piloted on three research databases (Business Source Complete, International Bibliography for the Social Sciences, and Scopus). We filtered the results of this pilot by scanning abstracts, summaries and sources according to a number of agreed relevance criteria, such as only including items written in English, those published after Kahn's (1990) seminal article on engagement was published, and those that included empirical data, generating a total of 5,295 items of literature. To achieve the inclusivity associated with the narrative synthesis approach, we conducted a separate pilot search for online grey literature adopting a somewhat less structured approach using search engines (e.g. Google) as well as embedded search tools from specific sources/sites recommended by the advisory group. This led to the identification of a further 80,000 items of grey literature.

Searching for and locating evidence

The volume and diverse nature of these results necessitated the production of a much more specific search approach, which was refined using the string: *'employee engagement' OR 'staff engagement' OR 'job engagement' OR 'organi* engagement' OR 'personal engagement' OR 'team engagement' OR 'psychological engagement' OR 'work* engagement' OR 'medical engagement'*. To ensure we did not omit important sources of literature, we included two further databases (Zetoc and Nexis) to minimize publication bias and to capture grey literature that could be searched in a more structured way (Patterson et al., 2007). This was enhanced by citation tracking, scanning reference lists, endnotes and footnotes as well as tracking alerts for new publications. Based on these refinements, our structured search produced a total of 7,932 items of literature from the five databases, which were imported into Refworks (Version 2.0). Using the Refworks 'de-duplication' functions brought this number to 5,771 items which then formed the basis of the next (sifting) stage of the review. An overview of the process is provided in Figure 1.

INSERT FIGURE 1 HERE

Sifting search results

To assess the materials for inclusion in the synthesis, these 5,771 items were first sifted independently by two members of the research team using a pro forma aimed at reducing bias based on a range of quality (e.g. 'opinion piece only – no empirical data'; Cronbach's alpha lower than 0.7) and relevance (e.g. 'not related to the research questions') ratings. We did not include any studies which did not report on primary research other than for background reference. After consultation within the research team and with the advisory group, we took the decision to exclude studies that drew on measures of engagement that had been subject to widespread

criticism in peer-reviewed journals concerning their face or content validity (Christian, Garza, & Slaughter, 2011; Little & Little, 2006). This included studies that drew on the Gallup Q12 measure of engagement (e.g. Jones & Harter, 2005).

To facilitate a systematic approach to this volume of literature, and with the six members of the research team based in different institutions and countries, it was necessary to commission the development of a bespoke relational database using Excel Professional Plus 2010 that enabled us to merge and import search results from different database sources based on their descriptive fields (abstract, authors, year, title, abstract, etc.). This database enabled us to systematically allocate items to research team members for sifting and coding according to the agreed criteria and to produce summary output data on these processes in order to aid transparency. After trialing the database and the criteria for functionality and consensus, (inter-rater kappa score = 0.75 indicating ‘substantial agreement’, Viera & Garrett, 2005: 361) the project team were then randomly allocated an equal share of the 5,771 items with each item assigned to two reviewers. Where there was disagreement on inclusion between reviewers, the item was assessed by a third reviewer from the team. Where there was any doubt about whether to include an item or not, we erred on the side of caution and put the item forward for further investigation. Using this systematic approach to sifting, a total of 5,557 items were excluded using the quality and relevance criteria, of which many (2,047) turned out to be duplicated items, along with eight items not in the English language.

INSERT TABLE 3 ABOUT HERE

This left 214 items for inclusion in the evidence synthesis, along with a further 14 items of grey literature which could not be assessed using the above approach (see Table 3) because they did

not contain primary or secondary data that could be extracted in any systematic way, or they did not provide substantive bases for any claims being made. With the agreement of the advisory group, these were removed from the review and were summarized in a separate report (Authors 2015a). Full text versions of all included materials were then retrieved and downloaded to a shared cloud-based folder.

INSERT TABLE 4 ABOUT HERE

Systematic data extraction and evaluation

With the guidance of the advisory group, a data extraction form was developed to enable the evaluation of the included items. To support this process, we developed a topic guide based on the criteria of *adequacy*, *sensitivity*, *relevance* and *robustness* that took account of different evidence types in order to assess the weight (quality and reliability) of the evidence during data extraction (see Table 4). Of the 214 included items, 38 were purely conceptual in nature and contained no empirical data; these were used as a point of reference to support the emerging narrative. Most of the 172 empirical items involved cross-sectional quantitative studies, but we also identified 22 longitudinal or time-lagged studies, four qualitative studies and one study using mixed methods, and in the synthesis we attributed greater weight to evidence arising from these more complex approaches.

To extract the data, each item was read in full by a member of the research team and we systematically tabulated the findings from each study relating to all the research questions, breaking the findings down by disaggregating antecedents and outcomes, and tracking the publication dates. We also noted in detail how engagement had been defined, measured and theorized in each item. Records were rechecked for error and bias by a team member assigned

responsibility for each research question. Of the included items, four were meta-analyses. To avoid duplication, we did not extract data from the meta-analyses but these were used for comparison purposes.

Narrative synthesis aims to develop a coherent narrative that summarizes and describes the evidence base (Popay et al., 2006). To assist with this, we undertook thematic analysis of the empirical items in order to identify and map out the underpinning conceptual bases of the evidence. We undertook a painstaking analysis of the ways in which engagement had been defined and measured in empirical research since 1990 and, through tabulation and comparison, critically assessed the weight of evidence relating to each definition/measure. Neither of these undertakings had hitherto been attempted in the literature. In other reviews, including Shuck (2011) for example, the different models and definitions of engagement have tended to be given equal weighting, whereas we found that the overwhelming weight of evidence in fact relied on a single definition, the Utrecht Work Engagement Scale (UWES; Schaufeli, Salanova, González-Romá & Bakker, 2002). By systematically mapping the underlying conceptual frameworks used in the engagement domain, we were able to identify a total of 48 different theories used to ‘explain’ engagement; again, this theoretical eclecticism had not previously been identified. However, the majority of studies used one theory, the job demands-resources (JD-R) model, that is closely associated with the UWES conceptualization of engagement (Schaufeli et al., 2002). Conversely, we noted that in 21 items no obvious theoretical framework was used, which suggested a degree of uncertainty within the literature concerning the basis of engagement’s relationship with other constructs. Overall, our analysis of the definitions, measures and theories used within the engagement domain pointed at the same time to the dominance of one particular, psychological perspective at the expense of others, yet also to a fragmentation of viewpoints,

including the emergence of an HRM discourse around engagement that had taken place in recent years.

Our second main area of focus was the evidence base concerning the link between engagement and outcome variables at the individual (e.g. morale, wellbeing, commitment) and performance levels (e.g. in-role, extra-role and counter-productive performance). After extracting the relevant data from the included items, we computed average correlations for each outcome measure in order to assess the size and direction of effects. We evaluated 47 studies concerning *individual outcomes* in which the main measures employed included commitment, burnout and turnover intention. There seemed to be convincing evidence to support the link between engagement and positive forms of work attitudes/wellbeing, such as job/life satisfaction. However, our review also revealed that the evidence was inconclusive when engagement was broken down into its constituent (i.e. lower order) facets/dimensions, or when longitudinal study designs were used. In one time-lagged study for example, Yalabik, Popaitoon, Chowne & Rayton (2013) raised doubts about the nature of the causal relationship between job satisfaction and engagement, stating it was unclear if job satisfaction was an antecedent or an outcome of engagement, or an intervening variable in relation to engagement and other factors. Through our analysis, we were able to evaluate both the *breadth* of evidence and the *weight* of evidence to highlight the distinction between correlation and causation.

We evaluated 42 studies that considered *performance outcomes* and categorized these either at the individual (in-role, extra-role, deviant behaviors) or the higher (team, organization, work unit) performance levels. We found moderate support for claims that engagement was positively related to individual task performance and organizational performance, but less support for an association with extra-role performance (e.g. citizenship behaviors) and deviant behaviors. By

systematically extracting and tabulating the data from these studies, we were able to evaluate the evidence on engagement and performance at a more nuanced level than had previously been achieved. Based on this, our analysis showed some evidence concerning a positive link between engagement and performance, but the volume and weight of the evidence were much less convincing than had been reported or suggested elsewhere given the reliance on cross-sectional studies (Halbesleben, 2010).

Evidence concerning interventions and *antecedents* of engagement was extracted from 155 studies. We grouped the findings under five headings: *individual psychological states*; *experienced job design-related factors* (such as task significance, variety, meaningfulness); *perceived leadership and management support/style*; *perceived organizational support/climate* (including perceptions of colleagues/team and psychological contract), and *specific organizational interventions or activities* (e.g. mindfulness training or other interventions aimed at enhancing individual coping strategies, resilience, or those seeking to enhance choice and autonomy in ways of working). Although the predominant JD-R framework would lead researchers to suppose that most of the extant evidence related to job-related factors such as job design, in fact, these were explored in fewer than half (42%) of studies that focused on antecedents. Moreover, it was noteworthy that even in time-lagged studies (Hakanen, Schaufeli & Ahola, 2008) there was at best only mixed evidence to support the claim that job resources foster work engagement. There was considerable weight of evidence relating to the other topic areas, notably individual psychological states, leadership, and perceived organizational factors such as support or the psychological contract. The review brought to light the paucity of studies focusing on analyzing particular interventions aimed at fostering high levels of engagement, which would

arguably be of most interest to practitioners. There were just nine of these studies and the findings from them were contradictory and inconclusive.

Synthesizing conclusions and dissemination

It is at this stage that the narrative synthesis approach is seen to provide flexibility by allowing reviewers to decide how best to synthesize findings, in a way that is ‘fit-for-purpose’ based on the review questions and the nature of the evidence itself, rather than using a pre-determined technique (Briner & Denyer, 2010: 356). Popay et al., (2006) state that the value of narrative synthesis lies in its capacity to go beyond simple summation and to map relationships in the extracted data both within and between studies. Narrative synthesis draws on different types of evidence – quantitative, qualitative, contextual - to achieve this, but the overall conclusions should be framed in narrative language (Popay et al., 2006). Based on our research questions concerning individual and performance outcomes and antecedents of engagement, we identified a number of core themes in the extracted data using a *thematic synthesis technique* which was related to key theoretical categories that emerged from the included conceptual evidence. Snilstveit, Oliver & Vojtkova (2012: 419) explain that this technique allows evidence to be synthesized through rigorous tabulation and constant comparison using a combination of aggregative, integrative and interpretive techniques, based on agreed categories. This helps to ensure that findings are relevant for research and practice and can be disseminated in meaningful ways.

Using this approach enabled us to show that *engagement as a psychological state* had become the orthodoxy of engagement theory in the academic field (Authors, 2015b). Through tracking publication dates, we found that this emphasis emerged from 2003 onwards when Guest (2014)

and others noted the explosion of interest in engagement, signifying a departure from Kahn's (1990) earlier focus on personal role engagement and a shift towards the construct of work engagement. Most studies meeting the quality and relevance thresholds for inclusion in our synthesis (86%) adopted this approach and used the UWES to measure and evaluate engagement. This is an important finding in relation to the shape of the engagement literature given that the construct validity of the UWES measure, its transferability across cultural contexts, and the underlying transactional assumptions of the predominant JD-R model have been questioned (Salanova & Schaufeli, 2013; Bargagliotti, 2012; Cole, Walter, Bedeian & O'Boyle, 2010).

The theme of *engagement as management practice* was based on a relatively small number of qualitative studies which view engagement as an intervention - *doing engagement* – rather than engagement as a psychological or attitudinal state, *being engaged* (Truss et al., 2014). Largely based on qualitative research, these studies reflect a more critical and recently emerging orientation in the evidence that had not hitherto been identified in previous reviews as a distinct strand of the engagement research narrative (Jenkins & Delbridge, 2013). Identifying and labeling this innovation in the engagement literature is important for two reasons. Firstly, it represents a very different standpoint concerning the construct of engagement that is more closely aligned with the interests of HRM scholars given the focus on involvement and participation, as compared with the more psychological and acontextual approach of the work engagement literature. Secondly, this approach offers more promise in terms of bridging the gap between academics and practitioners interested in engagement in light of the emphasis on organizational interventions and action.

The overall narrative that emerged from our review, therefore, was one of a broad and shifting landscape, densely populated by interesting but diffuse academic and practitioner studies and featuring many sound, and also some unsubstantiated claims about the beneficial associations between engagement and a range of other factors and outcomes. The evidence suggested we should proceed with some caution as the terrain is perhaps a little more limited and less stable than is suggested in some of the more normative literature. In the next section, we reflect on the process and conduct of our review, including our dissemination strategy, how we managed the challenges these presented, and we consider the strengths and limitations of the narrative synthesis approach.

Discussion: strengths and challenges of narrative synthesis

Strengths of narrative synthesis

Overall, we found that the narrative synthesis approach offered an accessible methodology since it is based on clear and purposive guidelines for systematically identifying, collecting, analyzing and evaluating evidence on a given topic. It is a systematized approach that draws on the narrative tradition to allow flexibility in the conduct of the review, but it goes beyond this in helping to produce a rigorous, replicable and critical synthesis of heterogeneous, complex and voluminous evidence bases that would not otherwise be possible using traditional forms of literature review. In this way, narrative synthesis can address the criticism often leveled at other approaches to review which adopt a purely positivist approach with limited application to social contexts (Hammersley, 2001). In drawing on its systematic heritage, narrative synthesis enables empirical findings on a topic to be assembled and reviewed as a body of work through a focus on the nature, quality, and quantity of evidence rather than adopting a narrow focus on a small

selection of literature as often characterizes traditional reviews, or focusing purely on quantitative studies as is the case for meta-analysis.

Narrative synthesis yields summaries of empirical evidence that are useful for other researchers in the field who want to understand the nature, quality and trajectory through time of the evidence base in order to focus their study designs on unanswered questions or areas of ambiguity. This is particularly true where the approach draws attention to important, emergent studies that have potential to reshape the field, such as Jenkins & Delbridge's (2013) research into approaches to engagement practice. The systematic approach means it is possible to evaluate and integrate a larger body of evidence than is the case either with traditional literature reviews or aggregative approaches to systematic review.

Narrative synthesis is guided by clear research questions that help to focus the selection and evaluation of evidence, as opposed to merely comprising a general review or addressing ad hoc questions that arise in response to findings. This enables a systematic and thorough approach to identifying and evaluating the evidence base in order to highlight imbalances and gaps. Through careful tabulation and comparison of the evidence, in the case of our synthesis of engagement for example, we were able to highlight an over-reliance on cross-sectional data and a paucity of studies that adopted longitudinal or complex methods that would lend greater credence to claims of causality.

Narrative synthesis additionally goes beyond summation however by seeking fresh insights from the evidence (Greenhalgh et al., 2005). Using this systematic approach, we were able to identify and synthesize strong themes in the evidence as well as exclude a considerable amount of literature that in general features prominently in search results and in the public domain but

which does not stand up to close scrutiny. We also were able to develop a critical summary of the temporal patterns in the evidence rather than simply providing static summation or aggregation. For example, we observed that while the concept of engagement originated in Kahn's (1990) participant observation study of how people choose to invest their preferred self in work (personal role engagement), the academic and practitioner literature soon departed from this and become dominated by the positive organizational scholarship literature that views engagement as an affective-emotional state, with the attendant issues outlined above. The trend towards identifying engagement as management practice has emerged as a much more recent development, helping to highlight the source of confusion that has emerged with the expansion of the concept since Kahn's (1990) original article.

In our narrative synthesis of engagement, the role of an advisory group was instrumental in facilitating multi-disciplinary and multi-stakeholder collaboration, particularly in developing the research questions and scoping potential sources of evidence from a range of different research disciplines and practical contexts such as the grey and practitioner literatures. These included parallel themes in the fields of psychology, business and management, sociology, philosophy, economics and health, which would have been difficult to span without the guidance and involvement of a range of external experts. Their involvement helped to achieve inclusivity of evidence emerging from other contexts and thereby build the overall credibility of the final synthesis. It helped to bridge the gap between academia and practice, to encourage a more interdisciplinary and reflexive approach which contributed to the objectivity of the review, and played a role in ensuring that the research team's own potential biases did not influence the choice of search terms or the choice of studies that were included (Rousseau & Barends, 2011).

For example, there were areas of evidence we would not have identified if we had maintained our initial narrow search strategy based on the more established databases and sources.

The advisory group also helped to plan a high impact dissemination strategy across different academic, policy-making and practitioner contexts. This included the production of a range of practitioner reports and guides; a series of podcasts for national employers' organizations; several conferences and workshops attended by representatives from different organizations and sectors, as well as a number of academic and technical outputs now in the public domain.

Issues and challenges of using narrative synthesis

Korhonen, Hakulinen-viitanen, Jylha & Holopainen (2013) make the point that the quality of any systematic review depends entirely on the quality of the included material. Therefore, in addition to the more general challenge of locating, sifting and evaluating data from large bodies of literature, the most distinctive challenge of narrative synthesis is managing the paradoxical claims to both precision and flexibility. No amount of procedure or guidance, however systematic or flexible, can fully prepare researchers or practitioners for the challenges posed by adopting a fully inclusive approach to evidence gathering, or the resources that this entails. Nor can it anticipate the kinds of judgments about quality or procedure that have to be made, particularly where the intent is to include grey literature. Rousseau et al. (2008) add that including unpublished studies is a strength of systematic reviews, such that the rationale for excluding material on any ground must be thoroughly explored.

Given the sheer volume of information available on the topic of engagement and the resources at our disposal, it was not possible for us to include all of the grey literature items (e.g. dissertations, practitioner publications and conference proceedings), which is a limitation of our study. This

highlights the practical and logistical difficulty of undertaking a synthesis on a topic like engagement, characterized by a large and diffuse evidence base. As Schaufeli (2014: 19) presciently commented, perhaps the most important challenge when approaching the issue of engagement is ‘where to draw the line’.

In attempting to be inclusive, we had few available technologies to manage the literature search and had to develop bespoke tools to do so because there were none available that enabled this volume of literature to be collected, integrated, sifted and analyzed by a team of researchers under the time pressures imposed by a funded project. It is therefore vital in constrained resource settings to fully assess the scale and nature of evidence and understand what resources are necessary to evaluate it before committing to this level of analysis. Armed even with our bespoke tool, a large project team and an advisory group, we were still forced to make difficult decisions about where to draw the line concerning what constituted ‘best evidence’ (Slavin, 1996).

To seek precision on this important point, we revisited the topic of what should constitute ‘evidence’ with the advisory group. Rycroft-Malone, Seers, Titchen, Harvey, Kitson et al. (2004: 82) argue that what distinguishes evidence from other types of information is *credibility* derived from the process of testing. They state: ‘A unifying theme in all definitions of evidence is that, however evidence is construed, it needs to be independently observed and verified’. From this perspective, ‘evidence’ therefore refers to information that, at the very least, has been subject to separate, objective scrutiny. In order to manage the large volume of material whose quality and relevance were difficult to determine at the outset and within our resource constraints, it was agreed - contrary to our initial intention - to focus on materials that had been subjected to peer-review as a proxy for quality. This decision effectively meant we had to exclude the grey literature from our review. However, with the support of the advisory group, we produced a

separate summary report of key materials from credible sources within the grey literature (Authors 2015a).

The quality and use of grey literature in evidence reviews appears to be an enduring issue. In a more recent analysis of using the grey literature in systematic reviews, Adams, Smart, & Huff (2016) explain that the unruly and rapidly increasing volume of grey literature requires more sophisticated ways of determining quality and relevance than those currently in use. They suggest one approach is to create ‘tiers’ of literature, similar to the idea of an evidence hierarchy but based on the criteria of *fitness for purpose* and *credibility*. Yet they, too, reach the conclusion that these criteria must be determined on a ‘review-by-review’ basis, suggesting as we found that there is no hard and fast rule about including and assessing grey literature.

In trying to maintain rigor, we additionally took the decision to exclude certain types of study within the main evidence review, including studies using measures such as the Gallup Q12 on quality (verification) grounds (e.g. Thorpe, Baqai, Witters, Harter, Agrawal et al., 2012); those reverse-scoring the Maslach Burnout Inventory (MBI) as a measure of engagement based on relevance grounds (unless engagement was also considered separately in the study); and studies that only examined the link between demographic factors (such as gender or race) as antecedents of engagement as these were not directly relevant to our research questions.

To assess the quality of included items, we developed a guide (see Table 4) based on criteria of *adequacy* (sufficiency of approach to meet research aims), *sensitivity* (specificity to uncover findings and ethical issues), *relevance* (appropriateness of method) and *rigor* (reliability/dependability), each of which were developed to address specific quality expectations of quantitative, qualitative, mixed methods and meta-analytic techniques. All these decisions

were taken on the basis of the research questions and with the guidance and input of the advisory group. We cannot be fully sure whether the decisions were correct, or whether they limited the final narrative. Nevertheless, when faced with large bodies of evidence, much of which is diffuse and lacking in structure, and limited resources, judgments have to be made about what to include and exclude based on criteria such as relevance, quality, credibility and legitimacy.

Recommendations for future research

Using narrative synthesis

We recommend that researchers considering using the narrative synthesis approach to evidence review, especially in relation to complex topics like engagement, need to be very clear on the likely scope and nature of the relevant literature prior to undertaking their project. Pre-project preparation should include some pilot studies involving the range of sources of evidence, including the grey literature, in order to develop plans for the management of what may well turn out to be very significant volumes of research findings. The resource and labor-intensive nature of narrative synthesis in HRM should not be underestimated. Narrative synthesis may nominally implies some similarity with a traditional literature review but it is in practice a much more complex and challenging process. In the end, despite our best efforts, we were unable to reconcile the issues of quality and comparability between the academic and grey literatures because the boundary between precision and flexibility seemed intractable, and it was here that we had to draw the line. Perhaps, with other review topics, it might be possible to develop uniform criteria for relevance, quality, credibility and legitimacy that span the academic and grey literatures; but reviewers would probably face similar challenges as we did if the topic were as mature or complex as engagement. If the grey literature is to have a place in evidence reviews, it is perhaps best used at the planning and development stage in scoping out review parameters and questions.

To ensure inclusivity of approach, we recommend that the research team involved in narrative synthesis represents a range of expertise from different disciplines and methodologies, as topics in HRM research tend to be conceptually complex. We found that the evidence in our review on engagement, although primarily quantitative, positivist and psychological, also spanned qualitative, critical and sociological domains, making inter-disciplinary expertise crucial to its success. We recommend that the conduct of evidence reviews is guided by an advisory group. We were unable to find any evaluation of their role in relation to this type of review, but our experience was that an advisory group made a substantial difference to review conduct, helping to widen the review scope so as to be inclusive, and assisting in developing criteria to ensure relevance and quality of evidence, as well as to facilitate review dissemination.

Although the process of evidence review is codified, there is very little guidance available to help determine how the synthesis stage of the review should develop, or even how best to determine what type of synthesis would be best suited to the evidence. It is unclear if it is even necessary to make this decision in advance or whether this is a judgment better left to later, when the nature of the evidence is better understood. In the case of our review, our approach to the synthesizing of the evidence emerged iteratively among the project team during the course of data extraction and write-up, as the nature and form of the evidence base became apparent. We recommend that researchers give consideration at the outset to how they might go about managing the synthesis stage in order to ensure that the previous steps in the process build towards this.

Conclusion

Narrative synthesis is a nuanced technique that would be useful to explore and evaluate evidence on other topics in HRM which are conceptually complex, notably those where there is a significant volume of evidence that could be synthesized to provide fresh insights, or where

research in the field has reached a level of maturity or divergence that would warrant a systematic review. Work-life balance for example is an important area of HRM research where the practices and perceived benefits of flexible working and its links to performance, commitment, goal setting and organizational outcomes are unclear but about which there has been many studies. Similarly, the emergence of new organizational forms makes greater evaluation of the strategic role and necessary competencies of HR managers in coping with conflicting demands for flexibility and security particularly suited to narrative synthesis in bridging multiple and complex topics of existing evidence leading to fresh insights. We can confirm that there is no shortage of potential evidence out there: the issue is how to evaluate it.

References

Authors (2015a) [Reference withheld to retain anonymity]

Authors (2015b) [Reference withheld to retain anonymity]

Adams, R., Smart, P. & Huff, A. (2016). Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies. *International Journal of Management Reviews*. Published online / open access: doi: 10.1111/ijmr.12102 [date accessed: 1st December 2016].

Alfes, K., Shantz, A., Truss, C. & Soane E. (2013). The link between perceived human resource management practices, engagement and employee behaviour: a moderated mediation model. *International Journal of Human Resource Management* 24(2): 330-351.

App, S. Merk, J. & Büttgen, M. (2012). Employer Branding: Sustainable HRM as a Competitive Advantage in the Market for High-Quality Employees. *Management Revue*, 23(3): 262-278.

Badger, D., Nursten, J. & Woodward, M. (2000). Should All Literature Reviews be Systematic?, *Evaluation & Research in Education*, 14(3-4): 220-230, doi: 10.1080/09500790008666974.

Bargagliotti, LA. (2012). Work engagement in nursing: A concept analysis. *Journal of Advanced Nursing*, 68(6): 1414-1428.

Barnet-Page, E. and Thomas, J. (2009). *Methods for the synthesis of qualitative research: a critical review*. ESRC National Centre for Research Methods: Working Paper Series (01/09). London: EPPI Centre/IOE

- Benzies, KM., Premji, S., Hayden, KA., and Serrett, K. (2006). State-of-the-evidence reviews: advantages and challenges of including grey literature. *Worldviews Evidenced-based Nursing*. 3(2):55-61.
- Boaz. A., Ashby, D., Denyer, D., Egan, M., Harden, M., Jones, D., Pawson, R. & Tranfield, D. (2006). A multitude of syntheses: a comparison of five approaches from diverse policy fields. *Evidence and Policy*, 2(4): 479-502.
- Boell, S. & Cecez-Kecmanovic, D. (2011). Are Systematic Reviews better, less biased and of higher quality? *European Conference on Information Systems (ECIS 2011)*. Paper 223. [Available from: <https://www.cebma.org/articles/>].
- Briner, R. (2011). Systematic Review and Evidence Synthesis as a Practice and Scholarship, in: Rousseau, D. (ed.) *Oxford Handbook of Evidence-based Management*. England: Oxford University Press.
- Briner, R. & Denyer, D. (2010). Systematic review and evidence synthesis as a practice and scholarship tool. In: Rousseau, D. (ed.) *Handbook of evidence-based management: Companies, classrooms, and research*. England: Oxford University Press; 2010. p.328-374. (<http://www.cebma.org/wp-content/uploads/Briner-Denyer-Systematic-Review-Evidence-Synthesis.pdf>; accessed online March 2015).
- Briner, R., Denyer, D. & Rousseau, D. (2009). Evidence-Based Management: Concept Cleanup Time? *Academy of Management Perspectives*, 23 (9): 19-32.
- Britten, N., Campbell, R. Pope, C., Donovan, J. Morgan, M. & Pill, R. (2002). Using meta ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research & Policy* 7 (4): 209–215
- Buchanan, D. & Bryman, A. (eds.) (2011). *The Sage Handbook of Organizational Research methods*. London: Sage.
- Campbell, R., Pound, M., Morgan, M., Daker-White, G., Britten, N., Pill, R., Yardley, L., Pope, C., & Donovan, J. (2011). Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. *NIHR Health Technology Assessment*. 15(43). doi: 10.3310/hta15430.
- Campbell Collaboration, (2001). *Guidelines for Preparation of Review Protocols*. Version 1.0, URL: www.campbellcollaboration.org/.../C2_Protocols_guidelines_v1.pdf [accessed 29 November 2013].
- Christian, M., Garza, A. & Slaughter, J. (2011). Work engagement: a quantitative review and test of its relations with task and contextual performance. *Personnel Psychology*, 64: 89–136.
- Cochrane. A. (1972) *Effectiveness and efficiency: random reflections on health services*. London: Royal Society of Medicine Press.

- Cole, M., Walter, F, Bedeian, A. & O'Boyle, EH. (2010). Job burnout and employee engagement: A meta-analytic examination of construct proliferation. *Journal of Management* 2012; 38(5): 1550-1581.
- Crawford, E., Lepine, J. & Rich B. (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology* 2010; 95(5): 834-848.
- Dawidowicz, P. (2010). *Literature reviews made easy. A quick guide to success*. Charlotte: Information Age Publishing.
- Denyer, D. & Tranfield, D. (2011). Producing a Systematic Review. In: Buchanan, D. & Bryman, A. (eds.) *The Sage Handbook of Organizational Research Methods*. London: Sage: 671-689.
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E, Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. & Sutton, AJ. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, 6(35): 1-13. doi:10.1186/1471-2288-6-35.
- Ecklund, E. & Scheitle, C. (2007). Religion among Academic Scientists: Distinctions, Disciplines, and Demographics. *Social Problems*, 54(2):289-307.
- EPPI Centre (2016). *Why is it important to be systematic?* [Online] Available at: <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=69> [Accessed 16th October 2016].
- Erez, A. & Grant, A. (2014). Separating Data From Intuition: Bringing Evidence Into the Management Classroom *Academy of Management Learning & Education*, 13(1): 104–119. doi.org/10.5465/amle.2013.0098.
- Fletcher, L., Robinson, D., Bailey, C., Alfes, K., Holmes, J., Madden, A., Buzzeo, J. & Currie, G. (2014). *Measuring employee engagement and interpreting survey results*. Institute for Employment Studies: Brighton see: <http://www.nhsemployers.org/~media/Employers/Documents/Retain%20and%20improve/Measuring%20employee%20engagement.pdf> (downloaded May 2015).
- Gough, D., Oliver, S. & Thomas, J. (2013) *Learning from Research: Systematic Reviews for Informing Policy Decisions: A Quick Guide*. A paper for the Alliance for Useful Evidence. London: Nesta (last downloaded June 2015 <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=88>).
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O. & Peacock, R. (2005). Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Social Science and Medicine* 2005; 61: 417-430
- Greenhalgh, T, Robert, G., Bate, P., Kyriakidou, O., MacFarlane, F. & Peacock., R. (2004). *How to Spread Good Ideas: a systematic review of the literature on diffusion, dissemination and sustainability of innovations in health service delivery and organisation*. Report for the National

Coordinating Centre for NHS Service Delivery and Organisation R&D (NCCSDO). London: HMSO 08/1201/038.

Guest, D. (2014). Employee engagement: fashionable fad or long-term fixture? In: Truss, C., Alfes, K., Delbridge, R., Shantz, A. & Soane, E. (eds.) *Employee Engagement in Theory and Practice*, London: Routledge; 221-235.

Hakanen, J., Schaufeli, W. & Ahola, K. (2008). The job demands–resources model: a three-year cross-lagged study of burnout, depression, commitment, and work engagement. *Work and Stress*, 22: 224–241.

Halbesleben J. (2010). A meta-analysis of work engagement: relationships with burnout, demands, resources, and consequences. In: Bakker AB, Leiter MP (eds.) *Work Engagement: A handbook of essential theory and research*. Hove: Psychology Press; p102-117.

Hammersley, J. (2001). On ‘Systematic’ Reviews of Research Literatures: a ‘narrative’ response to Evans & Benefield. *British Educational Research Journal*, 27(5): 543-554.

Higgins, J. & Green, S. eds., (2008). *Cochrane Handbook for Systematic Review of Interventions*. *Cochrane Book Series*. Chicester: Wiley-Blackwell.

Hu, Q. & Schaufeli, W. (2011). Job insecurity and remuneration in Chinese family-owned business workers. *Career Development International* 2011;16:6–19.
doi.org/10.1108/13620431111107784

Humphrey, S., Nahrgang, J. & Morgeson, F. (2007). Integrating Motivational, Social, and Contextual Work Design Features: A Meta-Analytic Summary and Theoretical Extension of the Work Design Literature. *Journal of Applied Psychology*, 92(5): 1332–1356.

Jenkins, S. & Delbridge, R. (2013). Context matters: Examining 'soft' and 'hard' approaches to employee engagement in two workplaces. *International Journal of Human Resource Management*, 24(14): 2670-2691.

Kahn, W. (1990). Psychological conditions of personal engagement and disengagement at work. *The Academy of Management Journal* 1990; 33(4): 692–724.

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. *Joint technical report*. Keele University. Technical report number: TR/SE-0401, 2004.

Kitson, AL., Athlin, A, Elliott, J. & Cant, M. (2014). What’s my line? A narrative review and synthesis of the literature on Registered Nurses’ communication behaviours between shifts. *Journal of Advanced Nursing* 70(6): 1228–1242. doi: 10.1111/jan.12321.

Korhonen, A., Hakulinen-viitanen, T., Jylha, V. & Holopainen, A.(2013). Meta-synthesis and evidence-based health care - a method for systematic review. *Scandinavian Journal of Caring Sciences*, 27(4): 1027-1034.

Little, B. & Little, P. (2006). Employee engagement: conceptual issues. *Journal of Organizational Culture, Communication and Conflict*, 10(1): 111–20.

- Marler, J. & Fisher, S. (2013). An evidence-based review of e-HRM and strategic human resource management. *Human Resource Management Review* 23: 18–36. doi:10.1016/j.hrmr.2012.06.002.
- Morrell, K. (2008). The narrative of ‘evidence-based’ management: a polemic. *Journal of Management Studies* 45(3): 613-635. doi: 10.1111/j.1467-6486.2007.00755.x
- Morrell, K. & Learmonth, M. (2015). Against evidence-based management, for management learning. *Academy of Management Learning & Education*, 14(4): pp: 520–533. doi.org/10.5465/amle.2014.0346
- Patterson, M., Rick, J., Wood, S., Carroll, C., Balain, S. & Booth, A. (2010). Systematic review of the links between human resource management practices and performance. *Health Technology Assessment* Vol. 14 (51). doi: 10.3310/hta14510.
- Pawson, R. (2006). *Mentoring relationships: an explanatory review*. ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 21.
- Pawson, R., Greenhalgh, T., Harvey, G. & Walshe, K. (2004). *Realist synthesis: an introduction*. Submitted to ESRC Research Methods Programme, University of Manchester, RMP Methods Paper 2/2004.
- Popay, J., Roberts, H., Sowden, A., Pettigrew, M., Arai, L., Rodgers, M., Britten, M., Roen, K. & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*. A Product from the ESRC Methods Programme, version 1.
- Purcell, J. (2014). Employee voice and engagement. In: Truss et al., (eds.) *Employee Engagement in Theory and Practice*. London: Routledge: 236-250.
- Rich, BL., Lepine JA. & Crawford, ER. (2010). Job engagement: antecedents and effects on job performance. *Academy of Management Journal* 53: 617–35. doi.org/10.5465/AMJ.2010.51468988
- Rousseau, D. & Barends, EG. (2011). Becoming and evidence-based HR practitioner. *Human Resource Management Journal* (Provocation Series Papers: HRM in the 21st Century), 21(3): 221-235. doi: 10.1111/j.1748-8583.2011.00173.x.
- Rousseau, DM., Manning, J. & Denyer, D. (2008). Evidence in management and organizational science: Assembling the field’s full weight of scientific knowledge through syntheses. *Annals of the Academy of Management*, 2(1): 475-515.
- Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A. & McCormack, B. (2004). What counts as evidence in evidence-based practice. *Journal of Advanced Nursing*, 47(1): 81–90.
- Sackett, DL., Rosenberg, W., Gray, JA., Haynes, RB. & Richardson, WS. (1996). Evidence Based Medicine: What It Is And What It Isn't: It's About Integrating Individual Clinical Expertise And The Best External Evidence. *British Medical Journal*, 312 (7023): 71-72.

- Saks, A. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, 21(7): 600-619.
- Salanova, M. & Schaufeli, W. (2008). A cross-national study of work engagement as a mediator between job resources and proactive behaviour. *The International Journal of Human Resource Management*, 19(1): 116-131.
- Schaufeli, W. (2014). What is Engagement? in Truss et al., (eds) *Employee Engagement in Theory and Practice*, London: Routledge.
- Schaufeli, W., Bakker, A. & Van Rhenen, W. (2009). How Changes In Job demands and resources predict burnout, work engagement, and sickness absenteeism. *Journal of Organizational Behavior*, 30: 893–917.
- Schaufeli, W. & Salanova, M. (2011). Work engagement: On how to better catch a slippery concept. *European Journal of Work and Organizational Psychology*, 20(1): 39-46. doi: 10.1080/1359432X.2010.515981.
- Schaufeli, W., Salanova, M., González-Romá, V. & Bakker, A. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies* 2002; 3(1): 71–92.
- Shuck, B. (2011). Integrative literature review: four emerging perspectives of employee engagement: an integrative literature review. *Human Resource Development Review*, 10: 304–328.
- Slavin, R. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48(1): 9-18.
- Snilstveit, B., Oliver, S. & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness*, 4(3), 409-429, doi: 10.1080/19439342.2012.710641.
- Thorpe, J., Baqai, W., Witters, D., Harter, J., Agrawal, S., Kanitkar, K. & Pappas, J. (2012). Workplace engagement and workers' compensation claims as predictors for patient safety. *Journal of Patient Safety*, 8(4): 194-201.
- Torraco, R. (2005). Writing integrative literature review: guidelines and examples. *Human Resource Development Review*, 4(3): 356-367.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Toward a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14: 207–222.
- Truss, C., Alfes, K., Delbridge, R., Shantz, A. & Soane, E. eds. (2014). *Employee Engagement in Theory and Practice*. London: Routledge.
- Viera, A. & Garrett, J. (2005). Understanding inter-observer agreement: The kappa statistic. (Family Medicine) *Research Series*, 37(5): 360-363.

- Vishwanath, V. & HakemZadeh, F. (2012), "Toward a theory of evidence based decision making". *Management Decision*, 50(5): 832 – 867. doi.org/10.1108/00251741211227546
- Watson, T. (2004). HRM and Critical Social Science Analysis. *Journal of Management Studies*, 41(3): 447-467. doi: 10.1111/j.1467-6486.2004.00440.x.
- Wefald, A. & Downey, R. (2009). Construct Dimensionality of Engagement and its Relation With Satisfaction. *Journal of Psychology*, 143(1), 91–111.
- Wilkinson, A. & Redman, T. (2013). *Contemporary Human Resource Management. Text and cases*. Pearson: London.
- Whittemore R & Knafl K (2006). The integrative review: updated methodology. *Journal of Advanced Nursing*, 52(5), 546–553.
- Yalabik, Z., Popaitoon, P., Chowne, J. & Rayton, B. (2013). Work engagement as a mediator between employee attitudes and outcomes. *International Journal of Human Resource Management*, **24**, pp. 2799–2823.

Table 1: Dimensions and issues in evidence synthesis

	Types of evidence synthesis				
	<i>Aggregative</i>	<i>Integrative</i>	<i>Interpretive</i>	<i>Explanatory</i>	<i>Narrative</i>
Types of evidence included	quantitative	quantitative and qualitative	qualitative	any	any
Types of sources allowed	published and unpublished studies	published and unpublished studies	published and unpublished studies	any	any
Generic / typical approach (with example)	meta-analytic e.g. Supramony's (2009) review of <i>HRM bundles and organizational performance</i>	meta-theoretic, e.g. Marler & Fisher's (2013) study of <i>e-HRM</i>	meta-ethnographic, e.g. Dixon-Woods et al.'s (2014) study of <i>access to services by vulnerable groups</i>	theory of change, e.g. Pawson et al.'s (2006) realist review of <i>mentoring</i>	inclusive; multi-method e.g. Kitson et al.'s (2014) review of <i>workplace communication</i>
Issues / questions addressed	- common in medical/ experimental research but also in management where topic is established, a single / narrow question can be asked and data can be aggregated / assimilated based on consistency and comparability	- appropriate for emerging or mature topics requiring articulation or greater clarity (e.g. new model) where formal but different data types exist, from experimental and non-experimental research	- also termed 'critical interpretive synthesis', includes different types of evidence with the aim of developing new theory based on critical evaluation of the evidence rather than testing existing theory	- based on a realist perspective, this approach is described as relevant where there is little concept clarity or where interventions and outcomes are complex and difficult to measure	- allows multiple questions to be asked about a mature topic where there is dissensus or divergence on meaning, context and future directions / research questions
Goal	validity; replicability	triangulation; multi-level understanding	thematic viability	plausible explanation of intervention (in)viability in context	plausible explanation of effects and contexts
Strengths	effects are added and bias minimized to produce greater validity	compensating for single study weakness; promotes inter-disciplinarity and insight	explores complex issues and meanings; offers critique	seeks insight rather than judgment; adapts to evidence types	allows for heterogeneity of evidence; flexibility
Weaknesses	a-contextual; narrow focus	epistemic inconsistency; lack of balance between data types	validity and comparability of source data; subjectivity	may not be of interest to decision-makers who seek 'answers'	difficult to manage scope of possible evidence

Table 2: Research objectives and review questions

Research objectives	Review questions	Specific research questions
To review and evaluate theory and practice relating to models of staff engagement	1. How has employee engagement been defined, modeled and operationalized within the academic literature?	1.1 How is employee engagement defined within the academic literature and in the health context?
		1.2 How has engagement been measured and evaluated within the academic literature?
		1.3 What theories are used to underpin models of engagement within the academic literature?
	2. What evidence is there that engagement is relevant for staff morale and performance?	2.1 What is the evidence that engagement is relevant for staff morale a) within the workforce in general b) within the context of health?
		2.2 What evidence is there that engagement is relevant for performance at the a) individual b) unit, team or group c) organizational or d) patient/client level either within the workforce in general or in the context of health?
	To produce a set of evidence-based outputs that help and guide NHS managers in fostering high levels of staff engagement	3. What approaches and interventions have the greatest potential to create and embed high levels of engagement within the NHS?
3.2 What evidence is there concerning approaches and interventions within an organizational setting at either a) the individual b) the unit, group or team or c) the organizational level that create and embed high levels of engagement within the health context?		
4. What tools and resources would be most useful to NHS managers in order to improve engagement?		4.1 What tools and resources are currently available for NHS managers?
		4.2 What tools and resources would NHS managers find useful?

Source: [Reference withheld to retain anonymity]

Table 3: Sources of grey literature and numbers of included items from each

Source of grey literature	Number of included items
CIPD	5
IES	3
Kenexa	3
The King's Fund	1
GSR (Government Social Research Service)	1
Referred from academic literature search	1
Total number of items	14

Source: [Reference withheld to retain anonymity]

Table 4: Guide for Assessing the quality of methods of research / evidence gathering

Criteria	Quantitative	Qualitative	Mixed (additional issues)	Meta-analysis
Adequacy (i.e. sufficiency of approach, etc. to meet research aims)	Of sample, research design, etc. Content Validity: does the measure actually measure what is claimed Predictive validity: does the study predict the required outcomes Construct validity: are the measures in the study distinct; not too high correlation between the constructs	Are the listed factors of local / wider context made clear in the research account Is there sufficient recognition of the impact of these and other factors in the research / analysis to enable judgment as to the likely accuracy of claims	Are both approaches sufficiently well configured; are they inter-related or parallel (i.e. answering same or different questions) Does the approaches complement each other (elaboration, enhancement, illustration; clarification: Greene et al 1989)	Does the analysis address a focused question; does the analysis address this in a coherent way Was the validity of included studies appraised Is the meta-analysis based on a reasonable number of studies Were unpublished studies included and controlled for
Sensitivity (and specificity, to uncover findings that fit or don't fit with hypotheses or RQs. It also relates to ethical issues, e.g. equality & diversity)	Of research design to context / engagement	Is the sample 'purposive' to the aims of the research: how were participants selected? Is lay / local knowledge included and given credibility Has the research been adapted / refined to meet contextual variables; is this clear / are lessons evident	Do mixed methods provide additional value, e.g.: Are the results from one method used to inform the other (i.e. triangulated) Do contradictions between findings lead to a reframing of RQs	Are inclusion / exclusion criteria sensitive to the current interests to justify extraction Were important studies omitted: why?
Relevance (i.e. is method appropriate; is it within project parameters; evidence of impact)	To Evidence Review RQs Have appropriate scales been used Have appropriate statistical analyses been employed Are statistical thresholds and fit criteria observed	To Evidence Review RQs Does the sample provide relevant data to the nature of the research and the context from which it is selected	To EE Evidence Review RQs Is rationale for mixed approach clear / justified	What question is being addressed by the meta-analysis; does it correspond to the current research question(s) Has the meta-analysis been cited / reviewed; has it led to a demonstrable change in policy / practice – depending on date of publication
Robustness / Rigor (i.e. replicability [including feasibility]; systematic nature of research conduct; dependability of measure)	Reliability - Does Cronbach Alpha (.7) support reliability - Tests for common method bias for cross-sectional data Generalizability: to what extent can findings be applied to settings other than those in which they were established?	Is there rigor in the approach to data analysis: i.e., does it follow from the research questions and sampling technique What is the 'situational' or 'contextual' representativeness of the research: is this explicit or must it be inferred by the evaluator Does the researcher(s) attempt to properly interpret the data in terms of research questions and context or is the data largely reported without analysis	Is this approach replicable / worth replicating Do the approaches detract from one another	Is the data-extraction and assessment process replicable Are all processes documented; how many people were involved in the data appraisal / extraction processes

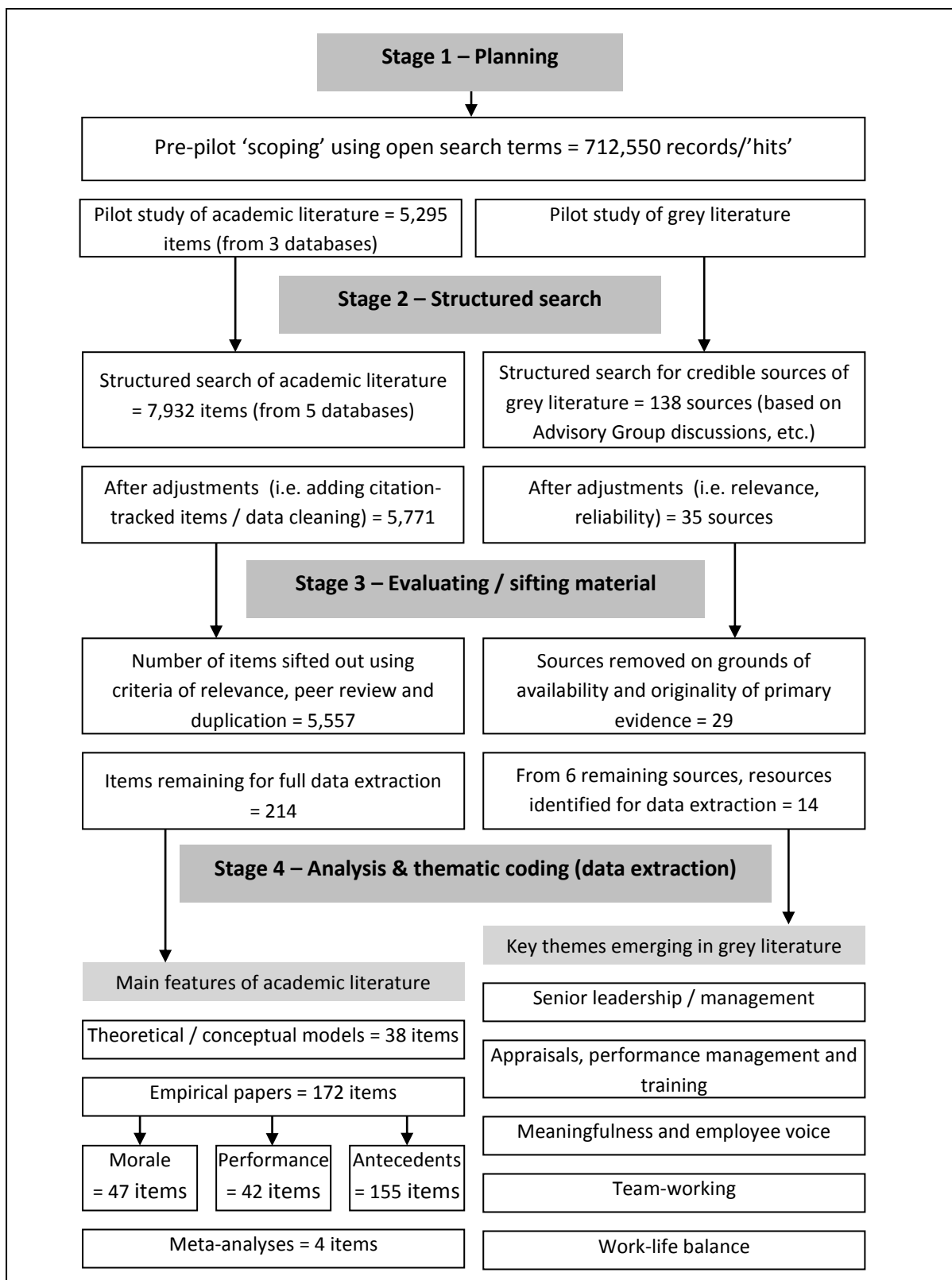


Figure 1: Prisma-style reporting flow-chart of stages 1 – 4 of narrative synthesis [Reference withheld to retain anonymity]

¹ The term ‘grey literature’ is seen to simply denote unpublished material, but it can often mean much more than this, including both informally published and unpublished material accessible through open source routes, such as government reports, committee reports’, as well as academic papers, theses, conference papers and proceedings and evaluations (Benzies, Premji, Hayden and Serrett, 2006: 56).