IJMR INTERNATIONAL JOURNAL OF MANAGEMENT REVIEWS
British Academy of Management
International Journal of Management Reviews, Vol. 19, 432–454 (2017)
DOI: 10.1111/ijmr.12102

Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies

Richard J. Adams, Palie Smart[1] and Anne Sigismund Huff[2]
Surrey University Business School, University of Surrey, Guildford, GU2 7XH, UK, [1]Cranfield School of Management, Cranfield University, Bedfordshire, MK43 0AL, UK, and [2]School of Business, Maynooth University, Maynooth, Co. Kildare, Ireland
Corresponding author email: correspond.adams@btinternet.com

This paper suggests how the 'grey literature', the diverse and heterogeneous body of material that is made public outside, and not subject to, traditional academic peer-review processes, can be used to increase the relevance and impact of management and organization studies (MOS). The authors clarify the possibilities by reviewing 140 systematic reviews published in academic and practitioner outlets to answer the following three questions: (i) Why is grey literature excluded from/included in systematic reviews in MOS? (ii) What types of grey material have been included in systematic reviews since guidelines for practice were first established in this discipline? (iii) How is the grey literature treated currently to advance management and organization scholarship and knowledge? This investigation updates previous guidelines for more inclusive systematic reviews that respond to criticisms of current review practices and the needs of evidence-based management.

Introduction

Summaries of past research are widely used both to inform new inquiries in many research disciplines and to influence professional practice (Briner *et al.* 2009; Shepperd *et al.* 2013). Procedures for systematic review and evidence-based decision-making based on review were developed for medicine by the Cochrane Collaboration (http://www.cochrane.org/) and received a good deal of attention from researchers from other fields in the early 1990s. A decade later, these ideas were adapted in the field of management and organization studies (MOS) (Denyer and Tranfield 2009; Rousseau 2006: 2012; Tranfield *et al.* 2003).

A critical feature of systematic review is comprehensive, rule-based search operations for collecting and synthesizing relevant evidence. Following the definition of a review question, identification of relevant knowledge is typically initiated by keyword/search string searches of electronic databases of scholarly publications. If these sources uncover a coherent body of high-quality, relevant, peer-reviewed articles, so-called 'white literature' (Lawrence *et al.* 2014), it is possible to proceed with scholarly

This work was conducted with support from the Engineering and Physical Sciences Research Council (EPSRC) Centre of Excellence for Industrial Sustainability grant number EP/I033351/1 (a research collaboration of Cambridge, Cranfield, Imperial and Loughborough Universities). No new data were created in the course of this work. The authors are grateful to Greg Boulton, Technology Enhanced Learning Designer at Cranfield University, for his help in creating Figure 1. The authors also acknowledge the pioneering work of Emeritus Professor David Tranfield who laid the foundations for systematic review in Management and Organizational Studies which continues to inspire.

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

inquiry on a firm foundation. However, scholars are increasingly recognizing instances where it seems appropriate to broaden the evidence search beyond the limits of academic journals to incorporate 'grey literature' (Adams *et al.* 2015; Sharma *et al.* 2015).

Incorporating grey literature – the diverse and heterogeneous body of material available outside, and not subject to, traditional academic peer-review processes – can make a variety of positive contributions to subsequent inquiry and practice. The review of reviews summarized in this paper shows that a significant number of MOS scholars already assert that grey material has relevance to their research questions and objectives. Our analysis reveals that these MOS scholars have used the grey literature to extend the scope of findings in their reviews by incorporating relevant contemporary material in dynamic and applied topic areas where scholarship lags; they have explored novel fields of enquiry, and have validated or corroborated findings from the academic literature.

Scholars active in other disciplines, with longer traditions of deploying systematic review, report additional benefits such as addressing publication bias (Hopewell *et al.* 2007), of which there is some evidence in MOS (Kepes *et al.* 2012), but limited in our sample. A number of reviewers from our sample who exclude the grey literature because of its challenges and the time required also believe that their conclusions may be poorer for its absence (e.g. Levy and Williams 2004). On the basis of these observations, we argue that there is strong justification for greater consideration of including the grey literature in future MOS systematic reviews.

While the potential contributions of grey literature are becoming apparent (Benzies *et al.* 2006; Rothstein and Hopewell 2009), little methodological guidance exists. For example, in the index of the 460 page *Oxford Handbook of Evidence-Based Management* (Rousseau 2012) 'grey literature' is referred to once. We believe that more specific guidelines for scholars on including grey literature in MOS reviews are important as the practice of systematic review in our field continues to mature. This paper contributes to discussion of the purpose and methods of literature review (e.g. Hart 1998; Jones and Gatrell 2014; Tranfield *et al.* 2003; Webster and Watson 2002) by suggesting how grey literature can be handled more systematically, even though its diversity also requires considerable flexibility.

Our purpose is to consider how current rules for MOS systematic review might be systematically broadened to incorporate grey literature, where 'grey literature' includes a wide variety of potentially relevant material, from specialist journals to blogs and other informal communications. Drawing on an analysis of 140 MOS systematic reviews published since Tranfield *et al.* (2003), we develop additional guidelines by considering three questions aimed at better understanding current use of this material:

1. Why is grey literature excluded from/included in systematic reviews in MOS?
2. What types of grey material have been included in systematic reviews since guidelines for practice were first established in this discipline?
3. How is the grey literature currently treated to advance management and organization scholarship and knowledge?

The analysis reported here more firmly connects systematic review to its original pragmatic and practice-oriented purpose (Tranfield *et al.* 2003) in the evidence-based management (EBMgt) paradigm (Briner *et al.* 2009; Huff *et al.* 2006; Rousseau 2012). A primary aim of the paper is to suggest further tools for pragmatic, 'evidence-informed' (Tranfield *et al.* 2003) research in MOS; 'pragmatic' in the sense that they inform future management practice by presenting a range of evaluated alternatives developed in the service of action (Pascal *et al.* 2013).

We begin by defining grey literature in relation to scientific/academic literature. Next, we describe our review methodology and present our synthesis of the evidence relating to our why, what and how questions about grey literature. Motivated by the need for improved management of the grey literature in systematic MOS reviews, our discussion draws our findings together in a set of practice guidelines, similar in intent to more general guidelines developed by Rousseau *et al.* (2008). A concluding discussion considers the use of grey literature in future reviews that take into account current critiques of systematic review practices.

Characterizations and challenges of grey literature

As a general definition, grey literature is composed of knowledge artefacts that are not the product of peer-review processes characterizing publication in scientific journals (Lawrence *et al.* 2014). Grey literature has been more specifically conceptualized in

Table 1. Instantiations of the grey literature[a]

Bibliographies	Rejected manuscripts	Publications from NGOs and consulting firms
Discussion papers	Un-submitted manuscripts	Videos
Newsletters	Conference abstracts	Wiki articles
PowerPoint presentations	Book chapters	Emails
Program evaluation reports	Personal correspondence	Blogs and social media
Technical notes	Newsletters	Data sets
Publications from governmental agencies	Informal communications	Committee reports
Reports to funding agencies	Census data	Working papers
Unpublished reports	Pre-prints	Company reports
Dissertations	Standards	Catalogues
Policy documents	Patents	Speeches
	Webinars	Reports on websites

[a]After Bellefontaine and Lee (2013), Benzies *et al.* (2006), Briner and Denyer (2012), Conn *et al.* (2003), Lawrence *et al.* (2014), Levin (2014), Petticrew and Roberts (2006), Rutter *et al.* (2010), Thomas (2008), Tyndall (2008).

narrow and broad ways. Schöpfel (2011) conservatively defines it as

> manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body.

Others offer more inclusive conceptualizations, often reflecting domain-specific idiosyncrasies (Bichteler 1991) that include an assortment of more ephemeral (re)sources unlikely to be systematically collected. Len Levin, librarian at Massachusetts University Medical School, defines grey literature as

> anything that has not been published in a traditional format or, in library parlance, lacks bibliographic control, meaning it can be hard to look up. This includes things such as conference proceedings, conference posters, dissertations and theses, government/institutional reports and raw data … luckily, much of it is now online … 'Institutional Repositories' … Government agencies – federal, state, provincial, etc. – … generate many reports that contain excellent data … [B]logs, Tweets or Facebook postings … can also be a great place to locate valuable information not found elsewhere. (Levin 2014)

Given these definitions, the types of material that constitute grey literature, as Table 1 illustrates, are diverse and come in a variety of forms.

Historically, access to many forms of grey literature has been a barrier to its inclusion in systematic review, since it has not been included in well-curated databases of academic disciplines. Through digitization, however, the size and influence of this type of literature has increased, and the need to include it in systematic review has become more evident, while better cataloguing and management has increasingly become a concern for librarians (Jeffery 2000). The fact that grey literature is not bound by the same publishing conventions that characterize white literature and comes in a variety of forms poses challenges for data management, extraction and synthesis. For example, grey literatures typically have no abstract, and so relevance and other inclusion criteria often cannot be determined without reviewing the entire document (Benzies *et al.* 2006). However, we anticipate that future advances in technology, including big data analytics and artificial intelligence, will result in significant improvement in the capacity to identify, access, review and incorporate grey literature (Lazer *et al.* 2009; Wilson *et al.* 2010).

Once potentially relevant grey literature is identified, questions of quality remain as another stumbling block to its inclusion in systematic review. Scholarly studies provide methodological descriptions that facilitate evaluating quality; such assurance is usually missing in grey publications, which tend to focus on conclusions rather than the process by which they were reached. Given that grey literature has rarely been through a peer review process and is generated in response to a much more diverse set of circumstances, there is a need to develop alternative and study-specific quality appraisal techniques and methods, where possible. Additional and less clearly defined standards add time and expense to the process of systematic review, which has deterred some researchers from using them (e.g. Conn *et al.* 2003), but it is reasonable to assume that high-quality work is published outside the white literature by individuals who are not under pressure to publish in academic

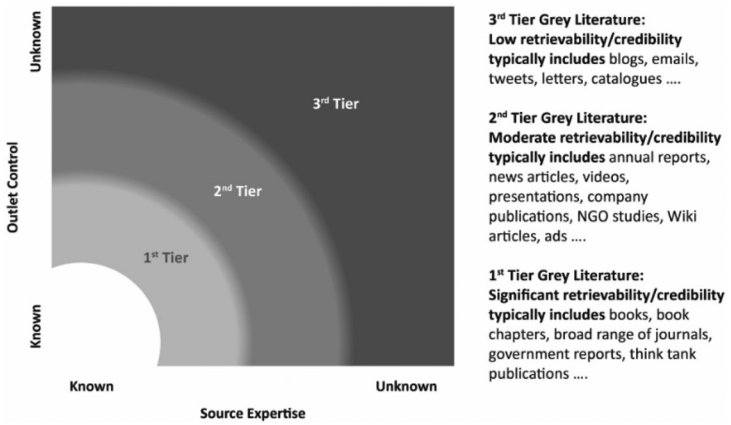

Figure 1. Shades of grey literatures

journals (Gibbons *et al.* 1994; Grayson and Gomersall 2003).

The heterogeneity of grey literature is a specific problem that makes it less amenable to traditional forms of archiving, retrieval, analysis, synthesis, bibliographic data capture, data extraction and integration. Conceptualizing around 'source availability', Kepes *et al.* (2012) propose a taxonomy of grey material that may be included in a meta-analytic review. Their four tiers describe sources from white to unidentifiable or unknown. Our observations of the range and types of grey literature included in and excluded from systematic reviews in MOS allow us to build on Kepes *et al.*'s (2012) taxonomy and posit gradations of grey literature using two dimensions (Figure 1). This gradation, shades of grey literature rather than discrete bands, is framed in terms of outlet control (the extent to which content is produced, moderated or edited in conformance with explicit and transparent knowledge creation criteria) and source expertise (the extent to which the authority of the producer of content can be determined).

This categorization recognizes that experts generate a range of material that may be of scholarly interest. Similarly, prominent outlets sometimes publish unreviewed material written by people with unknown training and experience. In the middle ground lie the many government reports, news articles, company publications and so on that may be of interest even though source expertise and outlet control cannot be fully determined. In all cases there are dangers of irrelevance, mistakes and fraudulent claims, as can be the case with the white literature (Barczak 2013; Cossette 2004), but in our opinion the more significant challenges of assessing grey sources require additional strategies.

Figure 1 emphasizes that boundaries between tiers tend to be fuzzy and permeable; therefore, the examples associated with each tier are only illustrative. Reviewers need to make explicit judgements about relevant grey literature on a project-by-project basis and reconsider categorization as sources and outlets for knowledge evolve. For example, the figure suggests that tweets are likely to be placed in Tier 3, since outlet and source are often unknown; however, in a closed conference of experts the source and 'publication' of tweets might more appropriately be classed as second or even first tier of grey evidence.

In short, the grey literature comes from a complex landscape of information artefacts generated in the course of real-life practices. Assessing this material requires time and complicated trade-offs whose appropriateness will vary according to study context. However, the grey literature can bring the disparate voices of experience into scholarly conversation to increase its relevance and impact.

Methodological approach

To identify current and potential ways in which grey literature might be used in MOS, systematic reviews published between 2003 (the first guidelines for systematic review in MOS published by Tranfield *et al.* 2003) and 2014 were targeted for analysis. Initially, 243 MOS journals were identified using the 2012 JCR Social Science Edition rankings of business and management journals, EBSCOHost Business Source Complete and also ISI Web of Science Social Sciences Citation Index (SSCI) Business and Management categories. Reviews from these journals were found by searching for the keywords 'systematic

Table 2. Outlets searched for systematic reviews published in the grey literature

Sponsor	Profile
NBS	Translates academic studies into practical planning resources that are relevant and actionable for Canadian business. The network commissions, among other things, systematic reviews on important sustainability issues with the goal of producing frame-breaking ideas that change management practice. www.nbs.net
EPPI-Centre	Located at the University of London but with an international mandate, the Centre has been undertaking, supporting and developing methods for systematic reviewing since 1993, principally in the fields of Education and Social Policy, Health Promotion and Public Health, International Health Systems and Development, and Participative Research and Policy. http://eppi.ioe.ac.uk/cms/
Campbell Collaboration	Helps people make well-informed decisions by preparing, maintaining and disseminating systematic reviews in education, crime and justice, social welfare and international development. Key user groups of reviews from each of these providers include practitioners and policy-makers. http://www.campbellcollaboration.org/

review' or 'systematic literature review' in the title, topic, abstract or subject terms. Articles were further filtered by reading their methodology section for an explicit statement and description of the systematic review process used, resulting in excluding 13 articles that claimed to be a systematic review, but failed to describe a recognizable systematic review methodology (e.g. Becker 2004; Carpenter *et al.* 2012). A further hand search of the *International Journal of Management Reviews* was undertaken to identify additional systematic reviews missed by the original search criteria. As a result of these steps, 124 systematic reviews published in the academic MOS literature were identified.

To understand the treatment of grey literature in systematic reviews published outside academic journals, a further search was undertaken to locate reviews in the grey literature. The sites of three reputable independent sponsors and advocates of systematic reviewing were searched for relevant reviews: the Network for Business Sustainability (NBS), the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) and the Campbell Collaboration. The three sponsors have somewhat different profiles (Table 2).

Following a hand search for systematic reviews relevant to the MOS domain from each of these websites, a total of 16 reviews (hereafter 'practitioner' reviews) were identified for analysis. In total, 140 systematic reviews (124 academic and 16 practitioner) comprise the sample for investigation.

The framework for analysing the use of grey literature draws on the generic principles and process of systematic reviewing (e.g. Jones and Gatrell 2014) and, more specifically, on Denyer and Tranfield's (2009) five-step outline, which focuses on: question formulation; locating studies; study selection and evaluation; analysis and synthesis, reporting and using results. Analysis was initiated by placing articles in one of three categories (Table 3).

Table 3. Categorization of sample articles

	Definition	Articles included
Category A	Articles in which grey literature is incorporated in review findings (see Appendix)	28 academic 16 practitioner
Category B	Articles in which grey literature is recognized as available, but is explicitly excluded	31 academic 0 practitioner
Category C	Articles that incorporate only the white literature and make no explicit reference to grey literature	65 academic 0 practitioner

Approximately 23% of the included academic reviews incorporate grey literature (Category A), approximately 48% acknowledge it as at least a potential source (Categories A and B), and 77% exclude it (Categories B and C). All 16 (100%) of the practitioner reviews include grey literature (Category A).

The origins of our research questions lie in our own experience of producing systematic reviews and working, in the context of EBMgt, with the resultant knowledge artefacts and with user communities and other stakeholder groups (Denyer and Tranfield 2009). Data and content analysis were facilitated using NVivo 10 for Windows (QSR International 2014). As shown in Figure 2, all Category A articles were analysed to help answer the 'why', 'what' and 'how' questions, Category B articles were considered for 'why' and 'what' questions and Category C to answer the 'why' question. Articles were coded deductively against this framework and then inductively analysed to allow for more nuanced understanding to emerge (Easterby-Smith *et al.* 2012) in each area.

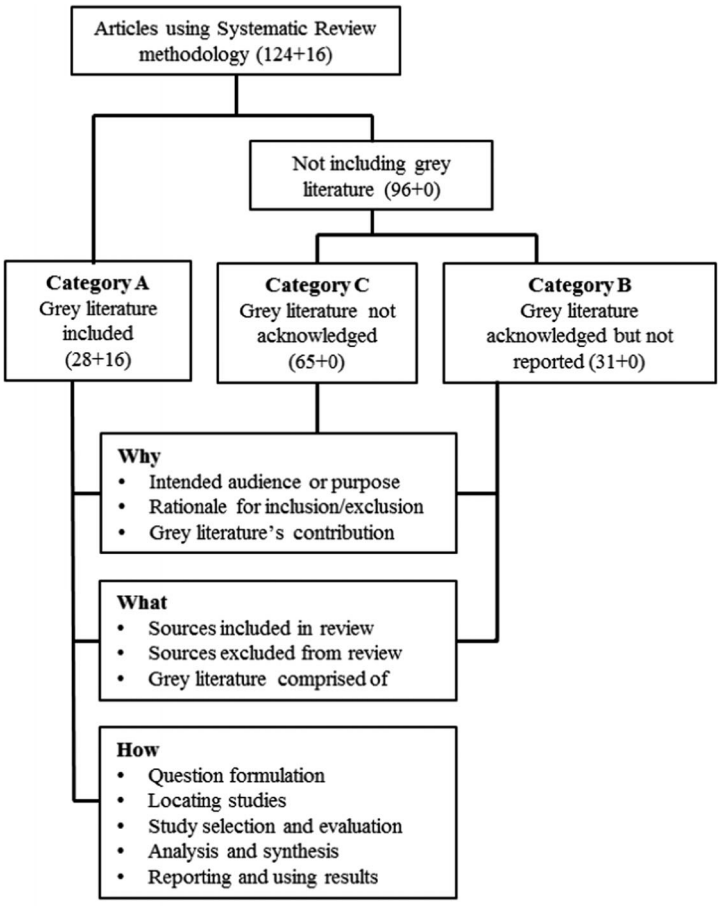

Figure 2. Search strategy and analytic framework

In the following discussion, we first reflect on the 'why' and 'what' questions of inclusion: what reasons – the advantages and benefits – were given for including the grey literature, and what material is considered? Subsequently, we turn to the 'how' question, of how scholars conduct reviews incorporating the grey literature. This means making decisions about when and when not to use the grey literature and decisions about its synthesis. A final section concludes with a presentation of findings addressing the question 'Why is grey literature excluded from systematic reviews?'

Findings

Why is grey literature included in systematic reviews?

Reviews in Category A (which include grey literature) often make explicit references to non-academic audiences as ultimate users of findings. Adams *et al.* (2012), for example, intend their findings to be used to stimulate discussion among both managers and senior executives on how firms might move toward sustainability. Albino *et al.*'s (2011) review is intended primarily for policy-makers and practitioners. The intended audience for Ton *et al.*'s (2013) report is people designing innovation grant systems for smallholders and those responsible for allocating funds. Brammer *et al.* (2011, p. 10) include 'significant practitioner knowledge in addition to the academic articles' in their review because of their intention to inform practice. These examples are drawn from practitioner reviews, but the user focus is also evident in academic reviews. Avard *et al.*'s (2010) review of public involvement in the area of human genomics, for example, is written to inform the public policy debate in the field by drawing on existing policy documents and guidelines.

Category A reviews were further analysed to establish the nature of grey literature's contribution to practice. Three reviews considering questions of measurement and metrics draw on grey literature to provide instances of measures and measurement practice not evident in the academic literature. Adams *et al.* (2006) looked for measures of innovation management in grey literature. Peloza and Yachnin (2008) found white and grey literatures reporting different types of (relevant) measurement: grey sources focused more than white on intermediate outcome metrics. Of the 20 environmental impact tools identified by Kaval (2011), 12 were identified in the white literature, one in the grey and the remainder across both literatures. Furthermore, within the respective literatures practitioners and academics differed in which tools were the focus of their attention: that is, different communities were interested in different things. Albino *et al.* (2011) turned to institutional reports to provide detailed data about emissions not available in academic studies. Neither Martin and Assenov (2012) nor Bowen *et al.* (2010) identify the specific contribution of grey literature in their reviews over and above that of the academic literature, but each article emphasizes that the added value of grey sources outweighs the negative effects of using a process that is not completely replicable.

Further benefits of grey literature are in its use not only to extend the range of evidence, but to fill gaps in the academic literature. Garengo *et al.* (2005), for instance, found a lack of evidence on performance measurement systems for small and medium-sized enterprises in the white literature and so turned to conference proceedings and internal reports and, in this way, identified new fields of inquiry. Corroborating this discovery opportunity, de Menezes and Kelliher (2011) report that a rich understanding of their substantive issue of interest was impaired by insufficient information about context provided by the academic literature, and supplement their review with grey literature. Similarly, Nitkin *et al.* (2009) can point to major adaptation initiatives in the insurance sector following a review of grey literature.

To complement and contextualize their findings on performance measurement Bititci *et al.* (2012, p. 307) explore general literature on global and business trends by drawing on the works of 'management gurus', and sources such as http://www.thinkers50.com and 'Who are the gurus' gurus?' The merit of this approach, they argue, is that it allowed for emerging conclusions from one stream of literature to inform the other. Perkmann and Walsh (2007) draw on complementary information from reports published by government agencies and other organizations (e.g. OECD) in their discussion of the UK's history of university–industry joint R&D establishments. Turner *et al.* (2013, p. 3) include grey literature for the reason that 'many of the major articles have been aimed at practitioners'; they intended the study to be as inclusive as possible and tried to avoid eliminating potentially valuable contributions.

This evidence leads us to the general conclusion that all tiers of the grey literature have the potential to define and contextualize phenomena of research knowledge because it includes potentially relevant knowledge that is sometimes not reported adequately in academic articles.

This review also shows that supplemental search of grey literature is undertaken for new insights when researchers feel their initial review strategy fails adequately to capture contextual information or address substantive concerns of the review. Some authors explicitly claim that the resulting review does not become any less 'systematic', and they show how they adhere to the tenets of transparency and rigour, arguably more so in light of the grey literature's heterogeneity.

Further, Adams *et al.* (2012), Albino *et al.* (2011) and Martin and Assenov (2012) all point to the contemporaneity or currency of grey literature, which can run ahead of academic research. Albino *et al.*'s (2011) study is a case in point: they draw on recent publications by industrial companies for evidence unavailable in the white literature. The practice of turning to grey literature for the most contemporary evidence dates back at least to Winn and Roome (1993, p. 148), who used sources ranging from company publications to journalistic articles in their effort to establish 'the state of the art of R&D management and the environment'.

Perceived gaps between researcher and practitioner interests are common also in Category A reviews. Where interests overlap, grey literature is typically used to support and validate findings from the academic literature. Nitkin *et al.* (2009) do this in their review of climate change risk in the agricultural sector, as do Adams *et al.* (2012) when cataloguing the activities associated with sustainability-oriented innovation.

We believe that the future of knowledge creation and thus systematic review is likely to be more interactive in this way, and suggest more specifically that grey literature be used when attempting to (a) validate/corroborate scientific outcomes, or (b)

Table 4. The grey literature included in and excluded from systematic reviews in MOS

Grey material included	Grey material excluded
Annual reports, blogs, business press, case studies, commercial organizations reports, commissioned reports, community engagement toolkits, conference proceedings/papers, consultancy reports, discussion papers, economic impact studies, editorials, government departments reports, government reports, industry reports, institutional reports, international organization industry reports, key works of few 'gurus', newspaper/magazine articles, NGO reports, patient opinions, policy documents, policy-maker consultations, practitioner articles, papers and reports, business press reports, research reports, teaching cases, theses (Bachelor, Honours, Masters, and Doctoral), think-tank reports, websites and working papers (Category A)	Anonymous publications, articles from popular rather than academic sources, book chapters, book reviews, books, commentaries practitioner accounts, conference papers, dissertations, editorials, essays, letters, journalistic or anecdotal articles, literature reviews, monographs, news items, non-refereed publications, opinion pieces, practitioner journals, prescriptive accounts, special issue introductory pieces, trade and popular press, unpublished papers, unpublished reports and working papers (Category B) Blogs, executive summaries of papers, newspaper articles, PowerPoint files, press releases, reports on the results of public consultations, reports or literature that describe implementation of consultation activities, short practitioner articles, unpublished posters, unsupported prescriptions and webmedia (Category A)

challenge assumptions/falsify results from academic sources.

What types of grey literature are excluded from/included in systematic reviews?

Table 1 illustrates the range of materials that systematic reviewers characterize as grey literature. Table 4 lends support to these instantiations by showing the materials identified as grey literature in Category A and Category B reviews, and portrays which were included/excluded in our sample studies.

The grey material that has been included in systematic reviews in MOS is clearly heterogeneously constituted. The most frequently cited grey material in the Category A reviews are conference proceedings and papers, (doctoral) theses and working papers. The least frequently cited include blogs, newspaper/magazine articles and the business press. The observation that Category A reviews deliberately reject some of the grey literatures indicates that selection decisions are not made indiscriminately, and active choices are pursued about inclusion and exclusion.

How is grey literature treated in systematic reviews?

Systematic review step 1: Question formulation. The purposes of Category B and C reviews are largely framed around academic preoccupations and concerns for scientific validity and advancement, as shown in Table 5. It is notable that Category C reviews do not consider grey literature, even though questions about impact and contribution in particular might be interpreted as benefiting from broader review, especially in dynamic or emergent fields where academic knowledge may lag.

As reported above, Category A reviews share some similar purposes to those outlined for B and C, but give greater prominence to non-academics as ultimate users of findings. Therefore, these reviews integrate user views as part of the stakeholder community (Denyer and Tranfield 2009) into the formulation of the review question. This is particularly appropriate where the research topic is novel or requires contextualization (Rutter *et al.* 2010).

Table 5 illustrates exemplars of research purposes and questions for Category A reviews that make explicit references to applied contexts, non-academic users/stakeholders and the pursuit of practical/policy impact. This table supports our earlier assertion – that the grey literature can contribute to research relevance – but adds the more specific idea that the grey literature can assist in question formulation when practitioner or user impact is a significant research purpose.

Systematic review step 2: Locate relevant studies. Locating studies for review requires a method by which studies relevant to a particular review question are found. Grey literature is tacitly excluded from many systematic reviews by dint of inclusion criteria restricting searches to well-established academic databases, reviewer decisions on search strings and keywords and frequently used search conventions (Denyer and Tranfield 2009). The decision to include grey literature thus requires conscious, explicit and different procedures, typically involving a review-specific strategy.

While replicability in subsequent studies may not be entirely possible when using grey literature

Table 5. Examples of research purpose

Explicit references to a context of application, non-academic users/stakeholders and practical/policy impact as purpose	Explicit references to a context of knowledge generation, academic users/stakeholders and theoretical impact as purpose
'To investigate the relationship between performance measurement systems and **SMEs'** (Garengo *et al.* 2005, p. 27)	'identify the degree to which the **marketing discipline** has hitherto engaged with **business model literature**' (Coombes and Nicholson 2013, p. 656)
'To better understand the theoretical content of Lean Product Development (LPD) research and potential reasons for **implementation** difficulties' (León and Farris 2011, p. 29)	'To establish both what we know and do not know about this topic, thereby **identifying areas for future research**' (Delbufalo 2012, p. 377)
'To examine the current research evidence on: the extent and nature of food promotion to **children** and, the effect, if any, that this promotion has on their food knowledge, preferences and behaviour' (Hastings *et al.* 2003, p. 1)	'This paper aims to provide a comprehensive understanding of the few studies conducted and **develop a detailed research agenda** to encourage **future research**' (Kauppi *et al.* 2013, p. 1368)
'To offer a systematic review of the literature that explores under-employment among recent **graduates''** (Scurry and Blenkinsopp 2011, p. 643)	'the authors analyze the field of international entrepreneurship (IE), which is in desperate need of **further theory development** … [They then] develop **directions for future research**' (Keupp and Gassmann 2009, p. 600)
'To help **managers** to understand what they can do to build and nurture a **sustainable** organization' (Adams *et al.* 2012, p. 10)	'This paper aims at **filling the gap** with a systematic literature review based on rigorous criteria' (Mari and Poggesi 2013, p. 171)
To understand how **policy** is made and how business can play a role in the development of sound **environmental policy**' (Auld *et al.* 2011, p. 5)	'. . . . [is] (IHRM) research is dominated by either a **universalist or contextualist paradigm**' (Sheehan *et al.* 2010, p. 1816)
''. . . we aim to collate extant academic and practitioner knowledge on **community engagement best practices**' (Bowen *et al.* 2008, p. 1)	'Proposes **a new direction for theoretical development** in this field' (Tatli and Özbilgin 2012, p. 180)
'. . . to synthesize the available literature in order to elaborate under what conditions innovation funds tend to be effective in facilitating innovation and **benefiting the poor and women in developing countries**' (Ton *et al.* 2013, p. vi)	(Exemplars from **Category B)**
'. . . summarizes the best available evidence for **business** to drive wider **social change** among the **public** . . . to identify evidence that relates **organizational actions to social change**' (Stephan *et al.* 2013, p. 9)	'How **do researchers define the concept of 'household'** in their **scientific papers**?' (Casimir and Tobi 2011, p. 498)
'To determine what evidence there is that the quality of initial **teacher education** (ITE) is influenced by its organisational structure, management processes' (Bills *et al.* 2007, p. 1)	'To understand whether the research . . . in the technology management (TM) field converge or diverge in terms of topics, **approaches, research focus, and methods**' (Cetindamar *et al.* 2009, p. 45)
'Whether **consumers** are willing to reward firms for their positive sustainability actions either by changing their behaviour or by paying a price premium' (Cotte and Trudel 2009, p. 6)	'To present recommendations on best practice **in conducting and publishing such research** derived from a Delphi study of authors who succeeded in doing so' (Fastoso and Whitelock 2011, p. 435)
'. . . to synthesize the current state of research on **business adaptation** to climate change, in order to identify and advance the theory and **practice** in this field' (Nitkin *et al.* 2009, p. 9)	'. . . to organize the research on trust between 2000 and 2011 . . . analyze the similarities and differences in **antecedents, consequences, and theoretical perspectives** dominant at each level' (Fulmer and Gelfand 2012, p. 1167)
'Systematic review of knowledge related to **managing sustainable supply chains** in an **international context**' (Brammer *et al.* 2011, p. 10)	'. . . a **focus on the methodological features** and developments over time' (Bergenholtz and Waldstrøm 2011, p. 540)
(Exemplars from **Category A)**	(Exemplars from **Category C**)

located with a unique strategy, systematic review requirements for transparency and traceability can be met by reviewers maintaining historical accounts via a researcher diary that records all searches. When extending searches beyond the electronic databases of published and unpublished studies (e.g. conference papers, working papers), a multilayered and semi-structured process is necessary to enable manageability, maintain control over a time-consuming process, and ensure transparency and replicability to the extent possible. Cataloguing the diversity of the search and specifying sources can be challenging. Some sites do not adhere to the traditional bibliographic format to which scholarly reviewers have become accustomed. However, maintaining a documentary record of decisions and processes enables researchers to report credibly for journal submission. Such records, if made publicly available, are useful for meeting the increasingly common stipulation of funding bodies such as Research Councils (UK) and the National Science Foundation (USA).

Researchers using grey literature typically search for it on technical and specialist online databases. Peloza and Yachnin (2008), for example, searched Greenbiz.com, Socialfunds.com, UN Environment Programme Finance Initiative, World Business Council for Sustainable Development and other grey sources. These sources were selected on the basis of their reputation, currency and authority as well as search functionality. Adams *et al.* (2012) explored a similar range of specialist sites as well as five blogs: authority and reputation their yardsticks for selection. Albino *et al.* (2011) screened 17 institutional databases, including those from consulting organizations, manufacturers associations, international research institutes and governmental organizations. Such practices were echoed across several Category A reviews (e.g. Avard *et al.* 2010; Stephan *et al.* 2013; Ton *et al.* 2013).

Many sites providing specialized knowledge are idiosyncratic in the way they present and archive knowledge content as well as in their search functionality. They are not equally easy to navigate, and often do not have sophisticated search facilities. More flexible processes for search are needed to explore both generalist and specialist sites. Consequently, and owing to the broad range of document type and search options, it may be necessary to modify or replace search strings developed for academic databases (e.g. Albino *et al.* 2011). In other words, it is important to explore electronic repositories (ideally with editorial independence and professional or institutional affiliations) to identify grey literature, creating site-specific search strings as necessary.

A number of reviews used generalized search engines such as Google, Google Books and Google Scholar (e.g. Bertels *et al.* 2010; Bowen *et al.* 2010; Brammer *et al.* 2011; Cotte and Trudel 2009; Kaval 2011). Such searches have necessitated developing new selection criteria in an endeavour to make the process as transparent as possible. In their Google searches, both Bowen *et al.* (2010) and Arvai *et al.* (2012) reviewed the first five pages of returns, and Bertels *et al.* (2010) the first 500 results. Kaval's (2011) Google search returned a nominal 462,000 results, but they found only 50 actual results visible and accessible. This experience raises interesting questions about internet searches, transparency and replicability. Palomino *et al.* (2013) demonstrate the instability of search engine results. They found that, rather than repeated search generating an increasing number of results as new content is added over time, results vary at an extremely fast pace – sometimes in only a matter of hours, with variable propensity of results to be higher or lower than in a previous search.

A more direct strategy used by reviewers, and less reliant on secondary sources, is to issue requests to practitioner and policy experts for relevant source literature (e.g. McDermott *et al.* 2006; Phelps *et al.* 2007). Experts queried may also be able to identify specialist websites with relative ease and professional confidence. Adams *et al.* (2012) considered this a useful generic approach, and made a similar request to scholars in the field of sustainability, but this strategy led to a disappointing response. Varied outcomes may be an indication of the immaturity of a field of inquiry and so further indicate a need for deeper exploration of grey literature. Therefore, we suggest deploying a semi-structured approach to identifying grey literature from generalist and specialist sites and augmenting results with sources identified by experts from policy and practice.

Systematic review step 3: Study selection and evaluation. It is necessary to specify the basis on which information sources have been included in or excluded from a systematic review before appraising the quality of included materials. An important and often overlooked point is that source evaluation is not simply a mechanism for excluding low-quality evidence, but is also about appraising and reporting what is included so that judgement can be made about the reliability of findings (Denyer and Tranfield 2009).

Evaluation is typically grounded in the research-quality debate. Systematic review practice in MOS draws heavily on the traditions of the medical domain, where the dominant evaluation paradigm is built around hierarchies of evidence (e.g. Evans 2003; OCEBM 2011), which privilege randomized controlled trials as the gold standard for evidence of what works (Rousseau *et al.* 2008). Under such conditions, a standardized approach to quality appraisal can be adopted. But, other standards apply when the included studies reflect different, diverse and sometimes unrevealed methodologies that characterize the heterogeneous grey literature.

A widely accepted tenet of systematic review is that individual articles are assessed for quality (Tranfield *et al.* 2003). One unexpected finding of the current study is the extent to which this has been diluted in MOS by application of a journal-level proxy alone. While there are a few exceptions (including Leseure *et al.* 2004; Pittaway *et al.* 2004; Reay *et al.* 2009),[1] the majority of published systematic reviews in MOS do not report evaluating quality at the level of the individual study.

Quality evaluation in MOS is difficult. The field has been characterized as ontologically and epistemologically diverse (Tranfield and Starkey 1998), making it difficult to develop a single and agreed paradigm for evaluation, the production of which has been described as 'a forlorn hope' (Johnson *et al.* 2006, p. 146). Consequently, multiple approaches to evaluation coexist. Gibbons *et al.* (1994), for example, distinguish between the quality considerations of Mode 1 and Mode 2 research, the latter being more socially accountable and reflexive. Alternative views on what quality means and how to assess it can be found in Boaz and Ashby (2003) and Starkey and Madan (2001).

Assessing quality is even more difficult when grey literature must be evaluated. Where it is reported, the principal evaluation criteria applied in Category A reviews are relevance and judgement (e.g. Gray and Stites 2013; Kiwanuka *et al.* 2011; León and Farris 2011). Bertels *et al.* (2010) ask, for example, 'Does the study examine antecedents of a sustainability culture?' and 'Does the study identify practices aimed at embedding sustainability?' Ton *et al.* (2013) were guided by an advisory board in identifying appropriate practitioner literature, but made selections independently. We extend this experience into three more general guidelines. First, it is often necessary to use fit-for-purpose quality criteria when selecting and evaluating grey literature. Second, it may be necessary to develop proxy measures of quality to sort large collections of literature, but justify them in a pilot exercise that considers the relevance and potential contribution of each artefact. Third, it often useful to be guided by field experts in evaluating grey literature, but retain decision-making independence and rationalize actions.

Systematic review step 4: Data analysis and synthesis. Analysis and synthesis is the breaking down of individual studies into constituent parts followed by reassembly, during which novel connections between the parts are made (Denyer and Tranfield 2009), all undertaken in a reflective and transparent manner. The diversity of the grey literature, and its distinctiveness from the white literature, raises new questions about whether or not and how more disparate evidence can meaningfully be synthesized.

Most forms of synthesis can be characterized as being either primarily interpretive or integrative, though each will often contain elements of the other (Dixon-Woods *et al.* 2004). Incorporating the grey literature tends to increase heterogeneity in the evidence base, but it remains accessible to multiple forms of synthesis, with selection dependent on the exigencies of each particular review. More specifically, Rousseau *et al.* (2008) identify four categories of purpose-oriented approach to synthesis, aggregative (e.g. meta-analysis), interpretive (e.g. meta-ethnography), integrative (e.g. content analysis) and explanatory (e.g. [critical] realist) approaches (see Bryman 2006; Lipsey and Wilson 2001; Noblit and Hare 1988; Pawson *et al.* 2004). In each of these forms of synthesis, the grey literature can have a role.

A relatively standard and accessible set of synthetic tools and procedures exist for reviewers to draw on. While some have been devised specifically for the purpose of synthesizing data from primary studies in mind, others are adaptions of techniques developed to analyse raw empirical data. Meta-analysis, as an example of the former, can be understood as a form of survey research in which research reports, rather than people, are surveyed (Lipsey and Wilson 2001); it requires homogeneous quantitative data.

Interpretive reviews can deal with diverse data. Reviewers tend to draw on techniques developed in the service of qualitative epistemologies, for

[1] Another exception is articles in journals that deal with management issues in medicine such as *Journal of Nursing Management*, where quality assessments of individual articles appear routinely.

instance thematic analysis (e.g. Thomas and Harden 2007), though they may also develop approaches inspired by meta-analysis such as qualitative meta-analysis (Schreiber *et al.* 1997; Zhao 1991) and meta-synthesis (Paterson *et al.* 2001).

In the Category A reviews, characterized by their inclusion of a diversity of grey literature in a variety of presentational formats and content, integrative and interpretive approaches to synthesis predominate. They principally deliver narrative-style reports using, though sometimes adapting, qualitatively oriented methods (e.g. variations in content and framework analytic and comparative approaches). Narrative summaries allow a highly flexible method, and reviewers appear to be taking advantage of this flexibility, though it appears they may lack awareness of alternative approaches.

The narrative mode of summary adopted in our sample resonates with Rousseau *et al.*'s (2008) 'synthesis by integration', involving the collection and comparison of evidence derived from two or more data collection methods in a search for patterns and connections. Given the increasing likelihood that data collection methods go unreported through increasing grey tiers of the literature (Figure 1), this can be understood as both a pragmatic and a compensatory strategy. It provides, first, opportunities for contextualization and validation – see, for example, Adams *et al.* (2012), who iterate between the academic and grey literature to validate a model of sustainability – and second, a compensatory mechanism to address individual study limitations through contextualization and triangulation.

In analysis and synthesis, it can be helpful to treat different categories of evidence separately. We believe, as already stated, that this approach should be used more often, because it preserves the unique qualities of each evidence type, provides evidence from each type that might be used to help interpret the other (akin to the 'reciprocal translation' of meta-ethnography discussed by Noblit and Hare 1988) and ensures that conclusions cannot be erroneously attributed to evidence sources (Mays *et al.* 2005). A small number of studies in our sample maintain such a distinction, thus enabling interpretations of findings that allow readers to determine the strength of inferences presented. McDermott *et al.* (2006) classified studies as of higher, medium or lower quality on a range of criteria, and reported findings against these qualifiers. Bertels *et al.* (2010) coded data at the level of 'sustainable practice', coding each practice that they uncovered in terms of level of empirical support. In addition to reporting empirically supported practices, they were able to include practices where practitioner knowledge led extant theory as well as instances where academics had proposed practices that had not been directly used or tested in practice. This approach is consistent with their objective to provide managers with a palette of sustainable practices to inform decision-making, and supports the view that grey literature can point to the wisdom of practice, which may not be reflected in scientific evidence (Benzies *et al.* 2006). It is also in line with Denyer and Tranfield's (2009) view that there is a greater likelihood that recommendations from systematic reviews in MOS will be heuristic rather than algorithmic.

In mature fields, where concepts and theories are clearer in advance, the grey literature enriches contextual understanding (e.g. de Menezes and Kelliher 2011; Lysaght *et al.* 2009). Pawson *et al.* (2004, 2005) argue that there is a critical role for the grey literature in 'realist' or 'theory-led' reviews where the aim of the project is primarily to develop theory for subsequent investigation and testing. This is a form of explanatory review in the Rousseau *et al.* (2008) categorization, and recognizes no hierarchy of evidence. Instead, as Gough and Elbourne (2002) have argued, the worth of the included study is determined only by the extent to which it contributes to discourse and pattern-building.

Emergent fields are characterized by inconsistency and proliferation of constructs, a widely distributed or fragmented body of knowledge, lack of a universal language, and absence of coalescing and binding theories (Burgess *et al.* 2006). The grey literature potentially provides additional evidence on which the processes of cumulation and generalization can act. In Rousseau *et al.*'s (2008) categorization, the purposes of this type of review are interpretation (building higher-order theoretical constructs) and integration (synthesizing across different methods to answer specific questions).

In a final observation on data collection and analysis, we find ourselves aligning with Hammersley (2005), who questions a prevailing positivistic dogma within the evidence-based practice community that privileges studies meeting some threshold of methodological rigour. Instead, we propose, where appropriate, the inclusion of grey literature not as a competing form of evidence, but as supplementary and complementary. In doing so, we argue that the findings from diverse bodies of evidence can be 'synthesized' to deliver value in terms of their contribution to discourse

and practice (Gough and Elbourne 2002). In sum, despite limited evidence in our sample of different approaches to synthesis being used, we contend that the grey literature can be used in explanatory synthesis to provide additional context specificity, in aggregative synthesis to counteract publication bias, in interpretive synthesis to offer a richer set of accounts and supplementary narratives, and in integrative synthesis to explore opportunities for triangulation and contextualization.

Systematic review step 5: Reporting and using findings. Denyer and Tranfield (2009) recommend a structured approach to reporting the process and findings of a systematic review that echoes frequently used standards of reporting empirical research. More specifically, the findings and discussion section of the final report should descriptively summarize the salient characteristics of all included studies. These will vary, depending on the nature of the review, but typically include the time period of publications, disciplinary origins, type of literature found (theoretical/empirical/conceptual), methodology, evolution of the field over time, description of method, industry sector and geographic domain.

Our analysis of Category A reviews shows that scholars consistently adopt this reporting practice but, as noted in the previous section, grey literature begins to lose its distinctive identity and contribution when treated and analysed as an equivalent to the white literature. For example, alongside searches of databases for the academic literature, de Menezes and Kelliher (2011) use Google to identify grey literature, including reports published by commercial or government organizations. Their selection criteria include relevance to the research question as well as theoretical and methodological rigour, but it is not clear whether these criteria were applied equally to both sets of potential studies in later phases of the review. Given that one of their included studies is an annotated bibliography, it seems unlikely that the same criteria could be applied to both white and grey literature identified. It is not our intention to be overly critical, but rather to illustrate a common tendency of current practice to conflate two separate bodies of literature with distinct characteristics and unique contributions. By failing to discriminate between what is and is not contributed from different sources, readers' confidence and willingness to use reported findings may be diminished.

Few Category A reviews suggest how findings from grey literature should be used alongside those from white literature. Of the few that do, Kaval (2011) discriminates between findings from the two as each assesses different types of tools. Peloza and Yachnin (2008) allude to differences in the type of literature reviewed, particularly with respect to perspectives on value in the context of sustainability. They note that results in the academic literature tend to show a less positive relationship between sustainability and financial performance than do grey reports. These studies are consistent with Rutter *et al.*'s (2010) suggestion that, while grey literature contains important and relevant insight and knowledge, summaries of such data should be reported separately from the synthesis of academic empirical studies so that readers and users are clear what level of data is informing which conclusions.

The overall discussion leads to the advice that report credibility for many audiences is likely to be enhanced through the inclusion of comparative and descriptive analysis of grey literature findings, especially in instances where findings supplement or challenge those of the white literature. Furthermore, unless academic and grey literatures are of similar status (as they may be in the case of reviews that include only Tier 1 grey literature), current use suggests that findings and confidence levels of systematic reviews of white and grey literature should be reported separately.

Why is grey literature excluded from systematic reviews?

Category C reviews, by definition, search only the academic literature and make no reference at all to grey literature. They account for just under 46% of the total of our sample reviews. These reviews primarily report that literature search was restricted to 'refereed international journals' (Van De Voorde *et al.* 2012), 'internationally esteemed journals' (Casimir and Tobi 2011; Fabbe-Costes and Jahre 2007), or 'peer-reviewed journal articles' (MacPherson and Holt 2007; Rose *et al.* 2011). The (often tacit) presumption in Category C reviews is that scholarly journals and, in particular, the double-blind peer-review process, are a reliable proxy for quality. A minority of authors make this assumption explicit and argue that the journal impact factor is an indicator of validated knowledge at the level of individual studies (e.g. Crossan and Apaydin 2010; Xuan *et al.* 2011). A small number delimit their searches still further by restricting searches to 'the most important journals', typically determined by league tables (e.g. Fulmer

and Gelfand 2012) or academic community opinion (e.g. Cetindamar *et al.* 2009).

Category B reviews draw only on the scholarly literature, and present a range of rationales for selection similar to those found in Category C reviews, but explicitly discount grey sources; a number explain why. Manageability of the review task is the critical concern, often in the face of an overwhelming volume of academic literature (e.g. Claus and Briscoe 2009; Leseure *et al.* 2004; Thorpe *et al.* 2005). In short, reviews are made more manageable by defining restrictive search criteria. However, it is important to note that the majority of Category B authors are unapologetic about excluding grey literature: they recognize that it exists, but consider publications from the academic literature adequate, at least for the purposes of the immediate project (e.g. Schmeisser 2013; Tarí 2011; Vishanth *et al.* 2009).

As with Category C articles that ignore grey literature, peer-review is used as a primary filter for inclusion as a proxy for quality in Category B articles. It is notable, however, that in a number of instances where authors reflect on the potential of grey literature, they also point to the loss of relevant input to the review, and the additional information that grey literature might contribute (e.g. Knoben and Oerlemans 2006; León and Farris 2011; Mari and Poggesi 2013). Claus and Briscoe (2009) more explicitly point to inherent biases in their selection criteria and say that additional insight could be gained from research outside their stated parameters. Knoben and Oerlemans (2006) describe their chosen method of searching literature as disadvantageous, leading to the omission of books and book chapters. Both Mari and Poggesi (2013) and Levy and William's (2004) reviews were limited by their self-confessed failure to exhaust the management literature because they excluded books, chapters in books, conference proceedings, working papers, dissertations and other unpublished works. Pittaway and Cope (2007) conclude that key aspects of their study were somewhat under-represented owing to their inclusion criteria and argue that future studies would benefit from examining grey literature in more detail.

In light of these findings, and in recognition of the fact that not all research purposes necessitate references to grey literature, we suggest a pragmatic approach for inclusion/exclusion decisions. For relatively mature and/or bounded academic conversations, with the possible exception of some Tier 1 literatures whose inclusion is defensible on the basis of established decision rules about quality, grey literature might be excluded. Where the grey literature potentially expands insight, but is judged too overwhelming to review, consider a pilot study or restrict scope in some other way to calibrate potential contribution further.

Discussion: systematic review, EBMgt and their critics

We have argued that one reason for including the grey literature in systematic reviews is to inform practice better. This raises the question of whether or not the findings and guidance produced in reviews that include the grey literature are in some sense 'better' for managers than they otherwise would have been had the grey literature not been included.

A number of the authors of Category A reviews explicitly state that they include the grey literature because they believe that, by doing so, they are able to provide better guidance for practice (e.g. Adams *et al.* 2012; Bertels *et al.* 2010; McDermott *et al.* 2006). It is tempting to infer that the guidance is *de facto* better for this inclusion, at least based on the professional view of the authors cited. At best, this is a tenuous rationale. The inference is strengthened, given evidence that some authors who do not include grey literature (e.g. Levy and Williams 2004; Mari and Poggesi 2013; Pittaway and Cope 2007) believe their studies would have been improved by its inclusion. But we lack robust, empirical, evaluative evidence from both the academic communities and the fields of practice that reviews are more impactful through the inclusion of the grey literature. Management and organization studies' collective experience with using systematic reviews has simply not progressed to an evaluation stage compared with fields with a more established history and practice. Even in more established fields with a longer history of including grey literature, the relationship between review, recommendations for practice and adoption by practitioners remains an important area of research (e.g. Cook *et al.* 1997; Owens 2011), one that MOS should be able to capitalize on in the future.

A range of evidence types is used in managerial decision-making (Briner *et al.* 2009; Kyratsis *et al.* 2014). However, metrics to test and juxtapose findings from systematic reviews and EBMgt are sparse, and more are needed (Reay *et al.* 2009). Future research needs to address such questions as whether or not and under what conditions the use of different evidence types leads to positive practice outcomes.

Further theoretical and experimental research is required to determine the nature of the contribution of the grey vs. white in different contexts, for example in established vs. emergent or dynamic vs. static theoretical contexts, according to the type of review question, and in terms of the review's ultimate purpose and use.

Two strategies are suggested to help address these questions. First, contrary to existing guidance on systematic reviewing in MOS (Rousseau 2012), previously published reviews in MOS have to a great extent been poor in methodological reporting. Of particular concern for the current paper is the relative failure to keep findings, conclusions and recommendations relating to white and grey literature separate. If in future these were kept distinct, the specific contributions of grey and white evidence would be easier to discern. Second, post hoc evaluations of managerial decision-making using different content are required. Once again, current knowledge is constrained by the quality of reporting. To our knowledge, no explicit and formal evaluations have been made of the contribution to practice of any Category A reviews. This is no surprise, given the relative immaturity of systematic review in MOS and uptake continues to remain a challenge in other more established domains of evidence-based practice (e.g. Gagliardi *et al* 2011; Shepperd *et al.* 2013).

Studies in other disciplines have suggested that incorporating the grey literature can also help address publication bias, the 'file drawer' problem (Hopewell *et al.* 2007). This exists when 'research that appears in the published literature is systematically unrepresentative of the population of completed studies' (Rothstein *et al.* 2005, p. 1), a tendency for journals to publish positive rather than weakly significant or neutral findings (Hopewell *et al.* 2007). The extent of publication bias in MOS is unclear, though one implication of its presence could be the promotion of practices based on potentially erroneous results (Kepes *et al.* 2012). One study included in our sample, a Category C review by Homberg and Bui (2013), reported that the published literature on the diversity performance link overestimates the strength of positive results. Some scholars within MOS have argued that the inclusion of the grey literature in systematic reviews can counterbalance such bias (e.g. Briner *et al.* 2009), though no Category A review reports such intent.

In MOS, the most likely effect of publication bias will be either to exaggerate effect or under-inform context by dismissing replication studies. Two things might contribute to why we have seen only limited activity in this space to date. First, although systematic review is no longer a novelty in MOS, since it is more than a decade since Tranfield *et al.*'s (2003) seminal paper, its practical application in the field has not reached maturity. Second, the same problems that afflict primary research, which include confirmatory and publication bias, may also afflict systematic reviews. That is, if the effect of including the grey literature is to diminish the strength of findings, might it not be more challenging to get these findings published too?

In future, we believe that MOS reviewers should pay greater attention to the possibilities and implications of publication bias, and respond with appropriate search and inclusion criteria. In particular, future reviews should make extensive efforts to locate all relevant studies and include those that show weak or negative findings to both test the contribution of grey literature and address possible problems of publication bias.

Evidence-based management

Several authors (e.g. Briner *et al.* 2009; Rousseau 2012; Tranfield *et al.* 2003) connect their calls for more systematic review to the applied and pragmatic intent of MOS and the emergent phenomenon of EBMgt. Briner *et al.* (2009, p. 19) specifically suggest that

> Evidence-based management is about making decisions through the conscientious, explicit, and judicious use of four sources of information: practitioner expertise and judgment, evidence from the local context, a critical evaluation of the best available research evidence, and the perspectives of those people who might be affected by the decision.

This paper shows how MOS researchers incorporate the grey literature in a fashion that reflects this diverse use of evidence. Our findings are consistent with Pawson *et al.* (2005), who argue that reviews based on the research literature alone can fail to provide a sufficiently rich, detailed and practical understanding of complex interventions. The grey literature has been shown in the Category A studies surveyed here to provide access to a wider variety of information than can be found in the academic literature alone. It provides a perspective for contextualizing, critiquing and reflecting on published studies, and so serves scholars and practitioners alike with data, knowledge and experience that can offer a more comprehensive and contextual view of the topic

of interest (Weintraub, n.d.). A principle criterion for inclusion of grey literature, then, becomes fitness for purpose, how it fits into the ways in which the findings are likely to be used (Boaz and Ashby 2003; Briner *et al.* 2009; Gough 2007; Nutley *et al.* 2013; Pawson 2006).

Evidence-based practice has captured the imagination of scholars and practitioners across a broad range of disciplines, including medicine, health care, education, public policy, social work and information science. This is illustrated by the NBS, EPPI-Centre and Campbell Collection websites searched for this review. In MOS, the EBMgt movement gathered momentum during the recent economic crisis, which rekindled the criticisms of management education articulated by Mintzberg (2004). Among other things, EBMgt has been seen as offering a new look at how to train the 'manager as practitioner' who can 'measure up to the standards not just of the academy, but also of particular professions' (Shulman 2005, p. 53; see also Burke and Rau 2010).

Systematic review can thus be seen not just as a first step in research and practice projects, but also as an important contribution to what can be accomplished in those projects. This is particularly important in dynamic and innovative environments where relatively little academic work has been accomplished and practice can be seen to be ahead of research investigations (e.g. Smart *et al.* 2007). We acknowledge that, as all complex conceptualizations in their infancy, both systematic review and EBMgt have to be further developed and tested (Briner *et al.* 2009). Management and organization studies is a relative latecomer to evidence-based practice (Madhavan and Niranjan 2015), but we are optimistic about its future as a promising sub-discipline in this effort precisely because MOS can benefit from previous experience. In other words, as previously stated, systematic reviews that include grey literature can provide an important source of information for research and practical projects, especially those concerning new fields of inquiry where knowledge from early experience is needed.

The recommendations for systematic review of grey literature developed in this paper also suggest how scientific evidence can be juxtaposed with other sources of evidence to provide a more pluralist stance for academic projects. Broader engagement of stakeholders is conducive to innovation in management practice because it increases the variety of interactions with evidence from past experience. This idea is in line with more general calls for increasing theoretic complexity with more diverse and pluralistic sources (e.g. Glynn and Dacin 2000).

Broad interrogation of evidence and its use is consistent with Pettigrew's call for 'a new concern for holism and sensitivity to action, dynamics, context and complexity'; he also suggests that scholarly assessment requires that variations in analysis are subject to knowing 'why might a particular course of action be the right thing to do in one situation and not another?' (Pettigrew 2011, p. 353). These ideas point toward the need to elicit what might be considered a 'just' course of action under certain circumstances. In such situations, the grey literature can draw on practitioner expertise and judgement from local contexts in which decisions, actions and outcomes occur and the perspectives of those people who might be affected by them. The opportunities offered by the grey literature thus are congruent with the sentiments of the EBMgt practising framework. We envisage positive repercussions for EBMgt as a practice that engages with multiple and discrete sources of evidence through systematic review and espouses the virtue of pluralism.

Continuing the discussion, we look more explicitly at those who characterize the pursuit of EBMgt as overly rationalistic and prescriptive. This concern is owed largely to a historical predilection for scientific evidence of only the highest quality (often ranked in hierarchies) as its input (Learmonth and Harding 2006), along with an observed tendency to pay more attention to the mechanics of review rather than the content of sources of material collected. In their critique of Tranfield *et al.* (2003), Morrell *et al.* (2015, p. 2) argue for 'equivalence between evidence and narrative or, rather, for recognizing narrative as evidence and evidence as narrative'.

In our view, systematic reviews that incorporate grey literature can and should accommodate a multiplicity of narratives that envelop the experiences and realities of practitioner and policy communities. When such narratives are juxtaposed and synthesized with other sources of evidence, they can empower managers and researchers to create their own meta-narratives regarding specific decisions or actions. In this way, systematic reviews that include grey literature support dialectical and dialogical propensities, and invite the generative possibilities of difference in a more democratizing manner (Bartunek and Rynes 2014; Depew and Lyne 2013; Romme *et al.* 2015;

Table 6. A decision aid for deciding whether or not to incorporate the grey literature in an MOS systematic review

Guideline 1 Including grey literature	Consider all tiers of grey literature to define and contextualize phenomena of research interest when potentially relevant knowledge is not reported adequately in academic articles, clearly stating the rationale for and source of material included in review.
Guideline 2 Including grey literature	Consider all tiers of grey literature when attempting to (a) validate/corroborate scientific outcomes with practical experience, or (b) challenge assumptions/falsify results from practice using academic research.
Guideline 3 Question formulation	Consult grey literature in formulating questions for research where practitioner or user impact is a significant purpose.
Guideline 4 Locating studies	Record and report all search decisions to support credibility of study process and findings and to build shared procedures for working with grey literature.
Guideline 5 Locating studies	Explore electronic repositories (ideally with editorial independence and professional or institutional affiliations) to identify grey literature, creating site-specific search strings as necessary.
Guideline 6 Locating studies	Deploy a semi-structured approach to identifying grey literature from generalist and specialist sites. Augment results with sources identified by experts from policy and practice.
Guideline 7 Selection and evaluation	Use fit-for-purpose quality criteria when selecting and evaluating grey literature. Develop proxy measures of quality to sort large collections of literature if necessary, but justify them in a pilot exercise that considers the relevance and potential contribution of each artefact.
Guideline 8 Selection and evaluation	Be guided by field experts in identifying sources for and evaluating grey literature, but retain decision-making independence and rationalize systematic review actions.
Guideline 9 Analysis and synthesis	Include the grey literature not as a competing form of evidence, but as supplementary and complementary evidence. Select a mode of analysis and synthesis consistent with the review question, nature of included evidence, and the intended purpose of the review.
Guideline 10 Using and reporting results	To increase report credibility for multiple audiences, consider comparative analyses, descriptive statistics, and inclusion of evidence that do not fit primary conclusions, with explanations of how decisions were made in analysis and synthesis.
Guideline 11Using and reporting results	Unless academic and grey literatures are of similar status (as they may be in the case of reviews that include only Tier 1 grey literature), findings and confidence levels of systematic reviews of white and grey literature should be reported separately.
Guideline 12 Excluding grey literature	Exclude grey literature from reviews supporting relatively mature and/or bounded academic conversations with the possible exception of some Tier 1 literatures that are relatively easy to defend on the basis of widely acknowledged decision rules about quality.

Trank 2014). This phenomenon is illustrated well by Trank (2014, p. 386) in her discussion about how the 'rhetoric of inquiry re-opens scientific texts to discussion' when performed by practitioners. From a meta perspective, engaging with grey literature in the systematic review process is thus attractive because it captures heterogeneity and promotes inclusiveness of the evidence pool and lessens the divide between stakeholders.

Guidelines for working with the grey literature in systematic review in MOS

As a summary of the review of systematic reviews summarized in this paper and related discussion, Table 6 presents a set of guidelines for working with the grey literature in systematic reviews in MOS.

Conclusion

This paper considers grey literature, broadly defined. Grey literature is produced by authors who may, but often do not, have academic training or interest in publishing in outlets that follow the norms of scholarly journals. We argue that expanding literature reviews to include purposefully selected material from the varied sources, though difficult, is increasingly important. More specifically, the variety, currency and utility of academic inquiry can be increased by considering the diverse wisdom gained in settings that concern MOS scholars.

Many scholars in MOS are responding to calls from research sponsors and evaluators to make their work more impactful and bridge the research-practice gap (Deadrick and Gibson 2007; Huff 2000; Starkey and Madan 2001). Some promising advances have been made in recent years (Hodgkinson and Rousseau 2009; Romme *et al.* 2015; Van de Ven 2007), and systematic review has been described as a means of further reducing discrepancies in knowledge and use by gathering and analysing evidence that can answer practice-relevant research questions (Tranfield *et al.* 2003). We believe that more inclusive reviews of

existing knowledge can expand many types of scholarly effort, including critical and literary projects.

There is an unruly and rapidly increasing amount of grey evidence available. A first contribution of this paper is to suggest that it be considered in tiers that begin with material similar to publications now widely accepted in academic reviews, then move outward to material that has great potential to add more novel insights, but is more challenging to assess in terms of source expertise and outlet oversight. While illustrative examples of grey material in each tier are offered, we suggest that relevance is determined and explicated on a review-by-review basis.

The second contribution of this paper is the discussion of grey material's potential contribution to research projects. The 12 guidelines that summarize this discussion are derived from a systematic review of reviews showing that grey material is already being considered and used by many MOS researchers, but then extended in light of trends, source availability and use in a digital age.

One requirement for assessing grey material more systematically is that quality determinations move beyond the status of journal publication as proxy, since, by definition, grey material is not closely controlled. A potential third contribution of the paper is, then, the suggestion that assessment of reviewed material be tailored to the research project as well as the source and nature of evidence considered. Given the amount of grey material that might be considered, decision rules for inclusion/exclusion are a particularly important part of this exercise. Pilot tests to justify decision rules are essential, as are collective efforts to establish shared norms in substantive areas of interest.

In Category A reviews that include grey literature, authors are clear that review findings would have been the poorer without drawing on ideas from grey sources. What is less clear is the extent, if any, of the impact on findings had the grey literature been included in Category B and C reviews. A number of authors reflected that their reviews might have been improved by the inclusion of the grey literature. Since many were concerned with the time required for review of more extensive sources, future researchers might consider running trials in which the same question is addressed, but by separate teams, one empowered to include the grey literature, the other not. Subsequently, the relevance of findings from a specific context of application might be determined by investigating user and stakeholder community perspectives and experiences. This signals a potential to design experimental studies concerning the adoption and diffusion of EBMgt and MOS ideas that might narrow the research–practice gap.

We align ourselves to the overall goal of more credible and generative management and organizational research. 'Credible' is a term used by those who consider research design (e.g. Robson 2011). It is connected to quantitative (Shipman 2014) and qualitative (Silverman 2011) methods and widely used in instructions for evaluating grey materials found on the web (e.g. http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html). The specifics and mechanics of how MOS research can become more credible is likely to evolve as discussion matures in a data-rich and digitally enabled context. Systematic reviews that include grey literature make a contribution to that endeavour.

References

Adams, R., Bessant, J. and Phelps, R. (2006). Innovation management measurement: a review. *International Journal of Management Reviews*, **8**, pp. 21–47.

Adams, R.J., Jeanrenaud, S., Bessant, J., Denyer, D. and Overy, P. (2015). Sustainability-oriented innovation: a systematic review. *International Journal of Management Reviews*, doi: 10.1111/ijmr.12068.

Adams, R.J., Jeanrenaud, S., Bessant, J., Overy, P. and Denyer, D. (2012). Sustainability oriented innovation: a systematic review. Available at: http://nbs.net/knowledge (accessed 1 May 2013).

Albino, V., Dangelico, R.M., Natalicchio, A. and Yazan, D.M. (2011). Alternative energy sources in cement manufacturing. A systematic review of the body of knowledge. Available at: http://nbs.net/knowledge (accessed 1 May 2013).

Arvai, J., Campbell-Arvai, V. and Steel, P. (2012). Decision-making for sustainability: a systematic review of the body of knowledge. Available at: http://nbs.net/knowledge (accessed 1 May 2013).

Auld, G., Burlica, B., Mallett, A., Nolan-Poupart, F. and Slater, R. (2011). When do climate policies work? A systematic review of experiences from low-carbon technology promotion and water management. Available at: http://nbs.net/knowledge (accessed 1 May 2013).

Avard, D., Stanton Jean, M., Grégoire, G. and Page, M. (2010). Public involvement in health genomics: the reality behind the policies. *International Journal of Consumer Studies*, **34**, pp. 508–524.

Barczak, G. (2013). Thoughts on academic research misconduct. *Journal of Product Innovation Management*, **30**, pp. 406–406.

Bartunek, J.M. and Rynes, S.L. (2014). Academics and practitioners are alike and unlike: the paradoxes of

academic–practitioner relationships. *Journal of Management*, **40**, pp. 1181–1201.

Becker, M.C. (2004). Organizational routines: a review of the literature. *Industrial and Corporate Change*, **13**, pp. 643–677.

Bellefontaine, S. and Lee, C. (2013). Between black and white: examining grey literature in meta-analyses of psychological research. *Journal of Child and Family Studies*, **23**, pp. 1378–1388.

Benzies, K.M., Premji, S., Hayden, K.A. and Serrett, K. (2006). State-of-the-evidence reviews: advantages and challenges of including grey literature. *Worldviews on Evidence-Based Nursing*, **3**, pp. 55–61.

Bergenholtz, C. and Waldstrøm, C. (2011). Inter-organizational network studies – a literature review. *Industry & Innovation*, **18**, pp. 539–562.

Bertels, S., Papania, L. and Papania, D. (2010). *Embedding Sustainability in Organizational Culture. A Systematic Review of the Body of Knowledge*. London, Ontario: Network for Business Sustainability.

Bichteler, J. (1991). Geologists and gray literature: access, use and problems. *Science and Technology Libraries*, **3**, pp. 39–50.

Bills, L., Briggs, M., Browne, A., Gillespie, H., Gordon, J., Husbands, C., Shreeve, A., Still, C. and Swatton, P. (2007). *Organisational Structures, Management Processes and Conceptual Framework in Initial Teacher Education: A Systematic Review*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Bititci, U., Garengo, P., Dörfler, V. and Nudurupati, S. (2012). Performance measurement: challenges for tomorrow. *International Journal of Management Reviews*, **14**, pp. 305–327.

Boaz, A. and Ashby, D. (2003). Fit for purpose? Assessing research quality for evidence based policy and practice. ESRC UK Centre for Evidence Based Policy and Practice, Working Paper 11.

Bowen, F., Newenham-Kahindi, A. and Herremans, I. (2008). *Engaging the Community: A Systematic Review. Knowledge Project Series*, Vol. **1**, Issue 1. London, Ontario: Research Network for Business Sustainability.

Bowen, F., Newenham-Kahindi, A. and Herremans, I. (2010). When suits meet roots: the antecedents and consequences of community engagement strategy. *Journal of Business Ethics*, **95**, pp. 297–318.

Brammer, S., Hoejmose, S. and Millington, A. (2011). *Managing Sustainable Global Supply Chains. A Systematic Review of the Body of Knowledge*. London, Ontario: Network for Business Sustainability.

Briner, R.B., Denyer, D. and Rousseau, D.M. (2009). Evidence-based management: concept cleanup time? *Academy of Management Perspectives*, **23**, pp. 19–32.

Briner, R.B. and Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. In Rousseau, D.M. (ed.), *The Oxford Handbook of Evidence-Based Management*. New York, NY: Oxford University Press.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*, **6**, pp. 97–113.

Burgess, K., Singh, P.J. and Koroglu, R. (2006). Supply chain management: a structured literature review and implications for future research. *International Journal of Operations & Production Management*, **26**, pp. 703–729.

Burke, L.A. and Rau, B. (2010). The research–teaching gap in management. *Academy of Management Learning & Education*, **9**, pp. 132–143.

Carpenter, M.A., Li, M. and Jiang, H. (2012). Social network research in organizational contexts: a systematic review of methodological issues and choices. *Journal of Management*, **38**, pp. 1328–1361.

Casimir, G.J. and Tobi, H. (2011). Defining and using the concept of household: a systematic review. *International Journal of Consumer Studies*, **35**, pp. 498–506.

Cetindamar, D., Wasti, S.N., Ansal, H. and Beyhan, B. (2009). Does technology management research diverge or converge in developing and developed countries? *Technovation*, **29**, pp. 45–58.

Claus, L. and Briscoe, D. (2009). Employee performance management across borders: a review of relevant academic literature. *International Journal of Management Reviews*, **11**, pp. 175–196.

Conn, V.S., Valentine, J.C., Cooper, H.M. and Rantz, M.J. (2003). Grey literature in meta-analyses. *Nursing Research*, **52**, pp. 256–261.

Cook, D.J., Greengold, N.L., Ellrodt, A.G. and Weingarten, S.R. (1997). The relation between systematic reviews and practice guidelines. *Annals of Internal Medicine*, **127**, pp. 210–216.

Coombes, P.H. and Nicholson, J.D. (2013). Business models and their relationship with marketing: a systematic literature review. *Industrial Marketing Management*, **42**, pp. 656–664.

Cossette, P. (2004). Research integrity: an exploratory survey of administrative science faculties. *Journal of Business Ethics*, **49**, pp. 213–234.

Cotte, J. and Trudel, R. (2009). Socially conscious consumerism. A systematic review of the body of knowledge. Available at: http://nbs.net/knowledge (accessed 1 May 2013).

Crossan, M.M. and Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: a systematic review of the literature. *Journal of Management Studies*, **47**, pp. 1154–1191.

De Menezes, L.M. and Kelliher, C. (2011). Flexible working and performance: a systematic review of the evidence for a business case. *International Journal of Management Reviews*, **13**, pp. 452–474.

Deadrick, D.L. and Gibson, P.A. (2007). An examination of the research–practice gap in HR: comparing topics of

interest to HR academics and HR professionals. *Human Resource Management Review*, **17**, pp. 131–139.

Delbufalo, E. (2012). Outcomes of inter-organizational trust in supply chain relationships: a systematic literature review and a meta-analysis of the empirical evidence. *Supply Chain Management – an International Journal*, **17**, pp. 377–402.

Denyer, D. and Tranfield, D. (2009). Producing a systematic review. In Buchanan, D. and Bryman, A. (eds), *The Sage Handbook of Organizational Research Methods*. London: Sage.

Depew, D.J. and Lyne, J. (2013). The productivity of scientific rhetoric. *Poroi*, **9**, p. 4.

Dixon-Woods, M., Agarwal, S., Young, B., Jones, D. and Sutton, A. (2004). *Integrative Approaches to Qualitative and Quantitative Evidence*. London: Health Development Agency.

Easterby-Smith, M., Thorpe, R. and Jackson, P. (2012). *Management Research*. London: Sage Publications.

Evans, D. (2003). Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, **12**, pp. 77–84.

Fabbe-Costes, N. and Jahre, M. (2007). Supply chain integration improves performance: the Emperor's new suit? *International Journal of Physical Distribution & Logistics Management*, **37**, pp. 835–855.

Fastoso, F. and Whitelock, J. (2011). Why is so little marketing research on Latin America published in high quality journals and what can we do about it? Lessons from a Delphi study of authors who have succeeded. *International Marketing Review*, **28**, pp. 435–449.

Fulmer, C.A. and Gelfand, M.J. (2012). At what level (and in whom) we trust: trust across multiple organizational levels. *Journal of Management*, **38**, pp. 1167–1230.

Gagliardi, A.R., Brouwers, M.C., Palda, V.A. and Lemieux-Charles, L. (2011). How can we improve guideline use? A conceptual framework of implementability. *Implementation Science*, **6**, p. 26.

Garengo, P., Biazzo, S. and Bititci, U.S. (2005). Performance measurement systems in SMEs: a review for a research agenda. *International Journal of Management Reviews*, **7**, pp. 25–47.

Gibbons, M., Limoges, L., Nowotny, H., Schwartman, S., Scott, P. and Trow, M. (1994). *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*. London, Sage.

Glynn, M.A. and Dacin, M.T. (2000). Pluralism and the problem of variety. *Academy of Management Review*, **25**, pp. 726–734.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, **22**, pp. 213–228.

Gough, D. and Elbourne, D. (2002). Systematic research synthesis to inform policy, practice and democratic debate. *Social Policy and Society*, **1**, pp. 225–236.

Gray, B. and Stites, J.P. (2013). Sustainability through partnerships: capitalizing on collaboration. Available at: http://nbs.net/knowledge (accessed 28 January 2014).

Grayson, L. and Gomersall, A. (2003). A difficult business: finding the evidence for social science reviews. *ESRC UK Centre for Evidence Based Policy and Practice, Working Paper 19*.

Hammersley, M. (2005). Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice. *Evidence & Policy*, **1**, pp. 85–100.

Hart, C. (1998). *Doing a Literature Review: Releasing the Social Science Research Imagination*. London: Sage.

Hastings, G.B., Stead, M., McDermott, L., Forsyth, A., Mackintosh, A.M., Rayner, M., Godfrey, G., Carahar, M. and Angus, K. (2003). *Review of Research on the Effects of Food Promotion to Children – Final Report and Appendices*. Prepared for the Food Standards Agency. Glasgow: Centre for Social Marketing, University of Strathclyde.

Hodgkinson, G.P. and Rousseau, D.M. (2009). Bridging the rigour–relevance gap in management research: it's already happening! *Journal of Management Studies*, **45**, pp. 534–546.

Homberg, F. and Bui, H.T.M. (2013). Top management team diversity: a systematic review. *Group & Organization Management*, **38**, pp. 455–479.

Hopewell, S., McDonald, S., Clarke, M. and Egger, M. (2007). Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*, doi: 10.1002/14651858.MR000010.pub3.

Huff, A., Tranfield, D. and Van Aken, J.E. (2006). Management as a design science mindful of art and surprise a conversation between Anne Huff, David Tranfield, and Joan Ernst van Aken. *Journal of Management Inquiry*, **15**, pp. 413–424.

Huff, A.S. (2000). 1999 Presidential address: changes in organizational knowledge production. *Academy of Management Review*, **25**, pp. 288–293.

Jeffery, K.G. (2000). An architecture for grey literature in a R&D context. *International Journal on Grey Literature*, **1**, pp. 64–72.

Johnson, P., Buehring, A., Cassell, C. and Symon, G. (2006). Evaluating qualitative management research: towards a contingent criteriology. *International Journal of Management Reviews*, **8**, pp. 131–156.

Jones, O. and Gatrell, C. (2014). Editorial: the future of writing and reviewing for *IJMR*. *International Journal of Management Reviews*, **16**, pp. 249–264.

Kauppi, K., Moxham, C. and Bamford, D. (2013). Should we try out for the major leagues? A call for research in sport operations management. *International Journal of Operations & Production Management*, **33**, pp. 1368–1399.

Kaval, P. (2011). *Measuring and Valuing Environmental Impacts. A Systematic Review of Existing Methodologies*. London, Ontario: Network for Business Sustainability.

Kepes, S., Banks, G.C., McDaniel, M. and Whetzel, D.L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, **15**, pp. 624–662.

Keupp, M.M. and Gassmann, O. (2009). The past and the future of international entrepreneurship: a review and suggestions for developing the field. *Journal of Management*, **35**, pp. 600–633.

Kiwanuka, S.N., Kinengyere, A.A., Rutebemberwa, E., Nalwadda, C., Ssengooba, F., Olico-Okui and Pariyo, G.W. (2011). *Dual Practice Regulatory Mechanisms in the Health Sector: A Systematic Review of Approaches and Implementation*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Knoben, J. and Oerlemans, L.A.G. (2006). Proximity and inter-organizational collaboration: a literature review. *International Journal of Management Reviews*, **8**, pp. 71–89.

Kyratsis, Y., Ahmad, R., Hatzaras, K., Iwami, M. and Holmes, A. (2014). Making sense of evidence in management decisions: the role of research-based knowledge on innovation adoption and implementation in healthcare. *Health Services and Delivery Research*, **2**(6).

Lawrence, A., Houghton, J., Thomas, J. and Weldon, P.R. (2014). *Where is the Evidence? Realising the Value of Grey Literature for Public Policy and Practice: A Discussion Paper*. Melbourne, Australia: Swinburne Institute for Social Research.

Lazer, D., Pentland, S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science*, **323**, pp. 721–723.

Learmonth, M. and Harding, N. (2006). Evidence-based management: the very idea. *Public Administration*, **84**, pp. 245–266.

León, H.C.M. and Farris, J.A. (2011). Lean product development research: current state and future directions. *Engineering Management Journal*, **23**, pp. 29–51.

Leseure, M.J., Bauer, J., Birdi, K., Neely, A. and Denyer, D. (2004). Adoption of promising practices: a systematic review of the evidence. *International Journal of Management Reviews*, **5/6**, pp. 169–190.

Levin, L. (2014). Literature search strategy week: Len Levin on understanding and finding grey literature. University of Massachusetts Medical School, Library Publications and Presentations.

Levy, P.E. and Williams, J.R. (2004). The social context of performance appraisal: a review and framework for the future. *Journal of Management*, **30**, pp. 881–905.

Lipsey, M.W. and Wilson, D. (2001). *Practical Meta-analysis*, Thousand Oaks, CA: Sage.

Lysaght, R., Donnelly, C. and Luong, D. (2009). Best practices in the rehabilitation of acute musculoskeletal disorders in workers with injuries: an integrative review and analysis of evolving trends. *Work*, **35**, pp. 319–333.

MacPherson, A. and Holt, R. (2007). Knowledge, learning and small firm growth: a systematic review of the evidence. *Research Policy*, **36**, pp. 172–192.

Madhavan, R. and Niranjan, A. (2015). *Evidence Based Management. Management and Leadership – A Guide for Clinical Professionals*. Heidelberg: Springer.

Mari, M. and Poggesi, S. (2013). Servicescape cues and customer behavior: a systematic literature review and research agenda. *Service Industries Journal*, **33**, pp. 171–199.

Mays, N., Pope, C. and Popay, J. (2005). Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research and Policy*, **10**, pp. 6–20.

Martin, S.A. and Assenov, I. (2012). The genesis of a new body of sport tourism literature: a systematic review of surf tourism research (1997–2011). *Journal of Sport & Tourism*, **17**, pp. 257–287.

McDermott, L., O'Sullivan, T., Stead, M. and Hastings, G. (2006). International food advertising, pester power and its effects. *International Journal of Advertising*, **25**, pp. 513–539.

Mintzberg, H. (2004). *Managers, not MBAs: A Hard Look at the Soft Practice of Managing and Management Development*. San Francisco, CA: Berrett-Koehler.

Morrell, K., Learmonth, M. and Heracleous, L. (2015). An archaeological critique of 'evidence-based management': one digression after another. *British Journal of Management*, **26**, pp. 529–543.

Nitkin, D., Foster, R. and Medalye, J. (2009). *Business Adaptation to Climate Change. A Systematic Review of the Literature*. London, Ontario: Research Network for Business Sustainability.

Noblit, G.W. and Hare, R.D. (1988). *Meta ethnography: Synthesizing Qualitative Studies*. Newbury Park, CA: Sage.

Nutley, S., Powell, A. and Davies, H. (2013). What counts as good evidence? Provocation paper for the alliance for useful evidence. Available at: http://www.alliance4usefulevidence.org (accessed 4 January 2014)/

OCEBM (2011). *The Oxford 2011 Levels of Evidence Table*. Oxford: Oxford Centre for Evidence Based Medicine. OCEBM Levels of Evidence Working Group.

Owens, D.K. (2011). Improving practice guidelines with patient-specific recommendations. *Annals of Internal Medicine*, **154**, pp. 638–639.

Palomino, M.A., Taylor, T., McBride, G. and Owen, R. (2013). Instability in search engine results: lessons learned in the context of horizon scanning applications. In Morvan, F., Min Tjoa, A. and Wagner, R.R. (eds), *24th International Workshop on Database and Expert Systems Applications (DEXA), 21–29 August, Prague*. Los Alamitos, CA: IEEE, pp. 53–57.

Pascal, A., Thomas, C. and Romme, A.G.L. (2013). Developing a human-centred and science-based approach to design: the knowledge management platform project. *British Journal of Management*, **24**, pp. 264–280.

Paterson, B.L., Thorne, S., Canam, C. and Jillings, C. (2001). *Meta-study of Qualitative Health Research: A Practical Guide to Meta-Analysis and Meta-Synthesis*. Thousand Oaks, CA: Sage.

Pawson, R. (2006). *Evidence-based Policy: A Realist Perspective*. London: Sage.

Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K. (2004). Realist synthesis: an introduction. *ESRC Research Methods Programme University of Manchester: RMP Methods Paper 2/2004*.

Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy*, **10**, pp. 21–34.

Peloza, J. and Yachnin, R. (2008). *Valuing Business Sustainability: A Systematic Review*. London, Ontario: Research Network for Business Sustainability.

Perkmann, M. and Walsh, K. (2007). University–industry relationships and open innovation: towards a research agenda. *International Journal of Management Reviews*, **9**, pp. 259–280.

Petticrew, M. and Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell.

Pettigrew, A.M. (2011). Scholarship with impact. *British Journal of Management*, **22**, pp. 347–354.

Phelps, R., Adams, R. and Bessant, J. (2007). Life cycles of growing organizations: a review with implications for knowledge and learning. *International Journal of Management Reviews*, **9**, pp. 1–30.

Pittaway, L. and Cope, J. (2007). Entrepreneurship education: a systematic review of the evidence. *International Small Business Journal*, **25**, pp. 479–510.

Pittaway, L., Robertson, M., Munir, K., Denyer, D. and Neely, A. (2004). Networking and innovation: a systematic review of the evidence. *International Journal of Management Reviews*, **5/6**, pp. 137–168.

QSR International (2014). *NVIVO 10 for Windows*. Available at: http://www.qsrinternational.com (accessed 16 May 2014).

Reay, T., Berta, W. and Kohn, M.K. (2009). What's the evidence on evidence-based management? *Academy of Management Perspectives*, **23**, pp. 5–18.

Robson, C. (2011). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*, 3rd edn. Chichester: John Wiley.

Romme, A.G.L., Avenier, M.J., Denyer, D., Hodgkinson, G.P., Pandza, K., Starkey, K. and Worren, N. (2015). Towards common ground and trading zones in management research and practice. *British Journal of Management*, **26**, pp. 544–559.

Rose, S., Hair, N. and Clark, M. (2011). Online customer experience: a review of the business-to-consumer online purchase context. *International Journal of Management Reviews*, **13**, pp. 24–39.

Rothstein, H.R. and Hopewell, S. (2009). Grey literature. In Cooper, H., Hedges, L.V. and Valentine, J.C. (eds), *The Handbook of Research Synthesis and Meta-analysis*. New York, NY: Russell Sage Foundation, pp. 103–126.

Rothstein, H.R., Sutton, A.J. and Borenstein, M. (2005). Publication bias in meta-analyses. In Rothstein, H.R., Sutton, A.J. and Borenstein, M. (eds), *Publication Bias in Meta analysis: Prevention, Assessment, and Adjustments*. Chichester: Wiley, pp. 1–7.

Rousseau, D., Manning, J. and Denyer, D. (2008). Evidence in management and organizational science: assembling the field's full weight of scientific knowledge through syntheses. *Academy of Management Annals*, **2**, pp. 475–515.

Rousseau, D.M. (2006). Is there such a thing as 'evidence-based management'. *Academy of Management Review*, **31**, pp. 256–269.

Rousseau, D.M. (ed.) (2012). *The Oxford Handbook of Evidence-Based Management*. New York, NY: Oxford University Press.

Rutter, D., Francis, J., Coren, E. and Fisher, M. (2010). *SCIE Systematic Research Reviews: Guidelines. SCIE Research Resource*. London: SCIE.

Schmeisser, B. (2013). A systematic review of literature on offshoring of value chain activities. *Journal of International Management*, **19**, pp. 390–406.

Schöpfel, J. (2011). Towards a Prague definition of grey literature. *The Grey Journal*, **7**, pp. 5–18.

Schreiber, R., Crooks, D. and Stern, P.N. (1997). Qualitative metaanalysis. In Morse, J.M. (ed.), *Completing a Qualitative Project: Details and Dialogue*. Thousand Oaks: Sage, pp. 311–326.

Scurry, T. and Blenkinsopp, J. (2011). Under-employment among recent graduates: a review of the literature. *Personnel Review*, **40**, pp. 643–659.

Sharma, T., Choudhury, M., Kaur, B., Naidoo, B., Garner, S., Littlejohns, P. and Staniszewska, S. (2015). Evidence informed decision making: the use of 'colloquial evidence' at NICE. *International Journal of Technology Assessment in Health Care*, **31**, pp. 138–46.

Sheehan, C., Fenwick, M. and Dowling, P.J. (2010). An investigation of paradigm choice in Australian international human resource management research. *International Journal of Human Resource Management*, **21**, pp. 1816–1836.

Shepperd, S., Adams, R., Hill, A., Garner, S. and Dopson, S. (2013). Challenges to using evidence from systematic reviews to stop ineffective practice: an interview study. *Journal of Health Services Research & Policy*, **18**, pp. 160–166.

Shipman, M.D. (2014). *The Limitations of Social Research*. Abingdon: Routledge.

Shulman, L.S. (2005). Signature pedagogies in the professions. *Daedalus*, **134**, pp. 52–59.

Silverman, D. (2011). *Interpreting Qualitative Data: A Guide to the Principles of Qualitative Research*. Los Angeles, CA: Sage.

Smart, P., Bessant, J. and Gupta, A. (2007). Towards technological rules for designing innovation networks: a dynamic capabilities view. *International Journal of Operations & Production Management*, **27**, pp. 1069–1092.

Starkey, K. and Madan, P. (2001). Bridging the relevance gap: aligning stakeholders in the future of management research. *British Journal of Management*, **12**(special issue), S3–S26.

Stephan, U., Patterson, M. and Kelly, C. (2013). Business-driven social change: a systematic review of the evidence. Available at: hhtp://nbs.net/knowledge (accessed 12 March 2014).

Tarí, J. (2011). Research into quality management and social responsibility. *Journal of Business Ethics*, **102**, pp. 623–638.

Tatli, A. and Özbilgin, M.F. (2012). An emic approach to intersectional study of diversity at work: a Bourdieuan framing. *International Journal of Management Reviews*, **14**, pp. 180–200.

Thomas, P. (2008). Shades of grey: the role of grey literature in systematic reviews. *PACEsetterS*, **5**, pp. 22–23.

Thomas, J. and Harden, A. (2007). *Methods for the Thematic Synthesis of Qualitative Research in Systematic Reviews. Evidence for Policy and Practice*. London: ESRC National Centre for Research Methods. Information and Co-ordinating (EPPI-) Centre, Social Science Research Unit.

Thorpe, R., Holt, R., MacPherson, A. and Pittaway, L. (2005). Using knowledge within small and medium-sized firms: a systematic review of the evidence. *International Journal of Management Reviews*, **7**, pp. 257–281.

Ton, G., de Grip, K., Klerkx, L., Rau, M.-L., Douma, M., Friis-Hansen, E., Triomphe, B., Waters-Bayer, A. and Wongtschowski, M. (2013). *Effectiveness of Innovation Grants to Smallholder Agricultural Producers: An Explorative Systematic Review*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Tranfield, D., Denyer, D. and Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, **14**, pp. 207–222.

Tranfield, D. and Starkey, K. (1998). The nature, social organization and promotion of management research: towards policy. *British Journal of Management*, **9**, pp. 341–353.

Trank, C.Q. (2014). 'Reading' evidence-based management: the possibilities of interpretation. *Academy of Management Learning & Education*, **13**, pp. 381–395.

Turner, N., Swart, J. and Maylor, H. (2013). Mechanisms for managing ambidexterity: a review and research agenda. *International Journal of Management Reviews*, **15**, pp. 317–332.

Tyndall, J. (2008). How low can you go? Towards a hierarchy of grey literature. Paper presented at Dreaming08: Australian Library and Information Association Biennial Conference, 2–5 September, Alice Springs.

Van de Ven, A.H. (2007). *Engaged Scholarship: A Guide for Organizational and Social Research*. Oxford: Oxford University Press.

Van De Voorde, K., Paauwe, J. and Van Veldhoven, M. (2012). Employee well-being and the HRM–organizational performance relationship: a review of quantitative studies. *International Journal of Management Reviews*, **14**, pp. 391–407.

Vishanth, W., Yogesh, K.D. and Zahir, I. (2009). The diffusion and use of institutional theory: a cross-disciplinary longitudinal literature survey. *Journal of Information Technology*, **24**, pp. 354–368.

Webster, J. and Watson, R.T. (2002). Analyzing the past to prepare for the future: writing a literature review. *Management Information Systems Quarterly*, **26**, p. 3.

Weintraub, I. (n.d.). The role of grey literature in the sciences. Available at: http://depthome.brooklyn.cuny.edu/library/access/greyliter.htm (accessed 13 2013).

Wilson, M.L., Kules, B., Schraefel, M.C. and Shneiderman, B. (2010). From keyword search to exploration: designing future search interfaces for the web. *Foundations and Trends in Web Science*, **2**, pp. 1–97.

Winn, S. and Roome, N. (1993). R&D management responses to the environment: current theory and implications to practice and research. *R&D Management*, **23**, pp. 147–160.

Xuan, Z., van Donk, D.P. and van der Vaart, T. (2011). Does ICT influence supply chain management and performance? *International Journal of Operations & Production Management*, **31**, pp. 1215–1247.

Zhao, S. (1991). Metatheory, metamethod, qualitative meta-analysis: what, why and how? *Sociological Perspectives*, **34**, pp. 377–390.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix: Summary table, Category A papers