

CEBMa Guideline for

# Rapid Reviews

in Management and Organizations



Version: 2.0

Editors: Eric Barends, Denise M. Rousseau, Rob B. Briner



## About this guideline

The CEBMa Guideline for Rapid Reviews in Management and Organizations is the official guideline that describes in detail the process of conducting a rapid review (RR) of the scientific literature on a managerial topic of interest. The aim of this guideline is to help authors to conduct an RR that meets the quality standards of the Center for Evidence Based Management (CEBMa). As such, it provides guidance on how to conduct an RR that is explicit about the question it poses, as well as being transparent, verifiable, and reproducible in terms of how the answers to this question are derived.

Although many guidelines are available on how to summarize scientific literature, most are written for specific areas, such as medicine, education or public policy. In addition, the majority of guidelines of this type are written for systematic reviews, rather than for rapid reviews. By providing a guideline that sets the appropriate standard for conducting an REA, we hope to contribute to a better understanding of the added value of RRs among managers and leaders, and to provide guidance to authors who aspire to conduct an RR on topics relevant to management and organizations.

## Disclaimer

The CEBMa Guideline is not a substitute for professional judgement and expertise. Deviation from the guideline, taking into account specific circumstances, may be appropriate. In addition, please note that this guideline is not a 'how to' handbook. For step by step instructions and in-depth information on how to conduct a systematic review or a rapid review we recommend Petticrew and Roberts' *Systematic Reviews in the Social Sciences: A Practical Guide* (Petticrew and Roberts, 2008) and David Gough's *An Introduction to Systematic Reviews* (Gough et al, 2012).

## Acknowledgements

The CEBMa Guideline for Rapid Reviews in Management and Organizations was developed by Eric Barends, Denise Rousseau and Rob Briner. However, most of the text is adapted from the '*Cochrane Rapid Reviews - Interim Guidance from the Cochrane Rapid Reviews Methods Group* (Garritty et al, March 2020). In addition, some of the text builds on articles, presentations, books and papers written by fellows of CEBMa. Finally, we are grateful to many of CEBMa's fellows who have given constructive feedback on the draft version of this guideline. We especially thank David Denyer, Michael Frese, Barbara Janssen, Richard MacKinnon, Pietro Marengo, Tony Munton, Claire Rahmatallah, Celine Rojon, Hanna Rothstein, Cédric Velghe, and Neil Walshe.

## How to cite this guideline

Barends, E., Rousseau, D.M. & Briner, R.B. (Eds). (2020). *CEBMa Guideline for Rapid Reviews in Management and Organizations*, Version 2.0. Center for Evidence Based Management, Amsterdam. Available from [www.cebma.org/guidelines/](http://www.cebma.org/guidelines/)

# Contents

Rapid Reviews	4
Requirements for reviewers	4
Steps in the RR process	5
Step 1. Background: What is the context of the RR question	6
Step 2. Formulating the RR question: What does the RR answer?	7
Step 3. Defining inclusion criteria: Which studies will be taken into account?	10
Step 4. Search strategy: How should the studies be sought?	11
Step 5. Study selection: How should you select the studies?	13
Step 6. Data extraction: What information should you extract?	15
Step 7. Critical appraisal: How to judge the quality of the studies.	17
Step 8. Results: What did you find?	21
Step 9. Synthesis: What does it all mean?	27
Step 10. Conclusion	31
Step 11. Limitations	32
Step 12. Implications of the findings for practice	33
Checklist	36
References	37
Appendix: Summarizing scientific literature	38

© CEBMa, 2020

Published by the Center for Evidence-Based Management, Amsterdam, The Netherlands

No rights reserved. All parts of this publication may be reproduced in any form or by any means without the prior permission of CEBMa. First published 2017

ISBN: 978-90-826957-0-0

For further information on CEBMa, visit our website: [www.cebma.org](http://www.cebma.org)

# Rapid Reviews

A Rapid Review (RR) provides a balanced assessment of what is known (and not known) in the scientific literature about an intervention, problem or practical issue by using a systematic review methodology to search and critically appraise empirical studies. However, in order to be 'rapid', an RR makes concessions in relation to the breadth, depth and comprehensiveness of the search. For example, an RR may exclude unpublished research, include only controlled studies, or limit the number of reviewers who critically appraise the studies' trustworthiness to only one. Due to these limitations, an RR is more prone to selection bias than a traditional systematic review. For a more detailed explanation of differences between a rapid review, a systematic review, and a traditional literature review, please see the Appendix: 'Summarizing Scientific Literature'.

## Requirements for reviewers

To successfully conduct an RR, reviewers should be able to:

1. find relevant search terms (keywords, related academic constructs, thesaurus terms, etc.);
2. determine the most relevant research databases for the RR question;
3. conduct a systematic and reproducible search in online research databases;
4. determine a study's methodological appropriateness and quality;
5. identify the impact (effect size) of a study;
6. assess and summarize a study's main limitations;
7. rate a study's trustworthiness;
8. assess and summarize a study's main findings, including practical relevance.

# Steps in the RR process

An RR involves the following steps:

1. Background
2. Question
3. Inclusion Criteria
4. Search Strategy
5. Study Selection
6. Data Extraction
7. Critical Appraisal
8. Results
  - 8.1. Definitions
  - 8.2. Causal Mechanism
  - 8.3. Main Findings
  - 8.4. Moderators and Mediators
9. Synthesis
10. Limitations
11. Conclusion
12. Implications for Practice



## Step 1. Background: What is the context of the RR question?

The background should clearly state what the rationale for the RR was and explain why the question being asked is important. You may also indicate how it might relate to a wider understanding of a general problem. Some RRs occur in the context of a specific organization. You should address this context (e.g. sector, history, characteristics), help specify the rationale for the RR, and explain why the question is important for the organization, its members or its clients. In formulating the RR question, it is important to reflect on the potential stakeholders relevant to the general problem being addressed in order to tap deeper insight into the issues involved (e.g., internal stakeholders such as employees at different organizational levels or external stakeholders like clients or the community).

### EXAMPLE 1

Despite the relevance of performance appraisals within the domain of human resource management, both academics and practitioners have had a somewhat uneasy relationship with them. Some academics question whether performance appraisals provide meaningful information, whereas others suggest that such reviews should be discontinued entirely. Given the widespread use of performance appraisals within our organization, we have conducted a Rapid Review to understand what is known in the scientific literature about the reliability and validity of performance appraisal and its impact on workplace performance.

### EXAMPLE 2

As a change consultant, I am expected to contribute to the realization of organizational change. The outcomes of change can be both positive and negative, depending on the type of change and the specific individual or group affected by the change. In both cases, but particularly when change has predominantly negative outcomes (i.e. lay-offs), I think it is of utmost importance that the change process is fair and just. I am curious which impact perceived fairness has on the way people perceive the outcomes of an organizational change.

### EXAMPLE 3

This RR was assigned to understand what academic research has discovered about the determinants of knowledge worker performance. For our company the pay-off for enhancing knowledge worker performance would be huge, not only in terms of financial outcomes, but also in terms of innovation. Although our company uses various measures and controls to monitor and enhance performance, we lack a basic understanding of what really drives knowledge worker performance.

## Step 2. Formulating the question: What does the RR answer?

You can use an RR to answer many different types of questions. For the purposes of this guideline, we split these into ‘impact’ and ‘non-impact’ questions. This distinction is not ideal but reflects the fact that the most common RR question are about

- the effect an intervention, factor, or independent variable.
- the drivers (antecedents) of a certain outcome.

### EXAMPLE 1

#### Main question

*What is known in the scientific literature about the impact of goal setting on the task performance of sales agents?*

#### Supplementary questions

- *What counts as goal setting?*
- *What is the assumed causal mechanism / how is it supposed to work?*
- *What is known about the overall effect of goal setting on task performance?*
- *What is known about possible moderators and/or mediators that affect the relationship between goal setting and task performance?*

### EXAMPLE 2

#### Main question

*What is known in the scientific literature about the drivers of unethical workplace behaviour and negligent misconduct?*

#### Supplementary questions

- *What is known about factors that may affect the incidence of unethical workplace behaviour and misconduct?*
- *How can these factors be measured in a valid and reliable way?*
- *What strategies for reducing unethical workplace behaviour and misconduct may be successful?*

## Non-impact questions

You can, however, use RRs to answer a range of other questions, which can be grouped as:

- **Needs:** What do people want or need?
- **Attitude:** What do people think or feel?
- **Experience:** What are peoples' experiences?
- **Prevalence:** How many / often do people / organizations ...?
- **Procedure:** How can we implement ...?
- **Process:** How does it work?
- **Exploration:** Why does it work?

### EXAMPLE

#### Main question

*What is known in the scientific literature about the prevalence of burnout among nurses in the US?*

#### Supplementary questions

- *What is burnout?*
- *What are the symptoms of burnout more widely and for nurses more specifically?*
- *Are there reliable and valid instruments available to measure burnout?*

## PICOC

A **PICOC** is a mnemonic used to assist reviewers to search for studies that are relevant to the professional context described in Step 1 (Background). The PICOC acronym stands for:

<b>P</b> opulation	Who?	Type of employee, subgroup, people who may be affected by the outcome
<b>I</b> ntervention	What or How?	Management technique/method, factor, independent variable
<b>C</b> omparison	Compared to what?	Alternative intervention, factor, variable
<b>O</b> utcome	What are you trying to accomplish / improve / change?	Purpose, objective, dependent variable
<b>C</b> ontext	In what kind of organization / circumstances?	Type of organization, sector, relevant contextual factors



Each element of the PICOC is vital in focusing the question and searching effectively for the most relevant evidence.

### EXAMPLE 1

**P:** physicians

**I:** 360-degree feedback

**C:** coaching

**O:** increased task performance

**C:** university hospital that has recently undergone significant organizational restructuring

### EXAMPLE 2

**P:** software developers

**I:** agile working

**C:** business as usual / status quo

**O:** reduced software development costs

**C:** large international IT firm in a highly competitive market

In addition, your PICOC will help you to determine whether the findings of a study will be generalizable and applicable to your organizational context. More specifically, your PICOC helps answer the question of whether your population, outcome of interest, and organizational characteristics are so different from those in the study that its results may be difficult to apply. After all, some psychological principles are generalizable to all human beings, but sometimes what works in one narrowly defined setting might not work in another.

## Step 3. Defining inclusion criteria: Which studies will be taken into account?

One of the features that distinguishes an RR from a traditional review is the pre-specification of criteria for including and excluding studies. The inclusion criteria (also referred to as eligibility criteria) help the reviewer(s) to determine whether a study will be included in the RR when reviewing its abstract and/or full text. The inclusion criteria should be guided by your RR question and objectives, and by the outcome measures that you will be considering to answer your question. They define the studies that the search strategy is attempting to locate.

### EXAMPLE

#### Inclusion criteria

1. **Date:** published in the period 1980 to 2020 for meta-analyses and the period 2000 to 2020 for primary studies
2. **Language:** articles in English
3. **Type of studies:** quantitative, empirical studies.
4. **Study design:** controlled studies
5. **Measurement:** studies in which the effect of goal setting on task performance was quantitatively measured
6. **Outcome:** task performance
7. **Context:** studies related to workplace settings.

#### Exclusion criteria

1. Studies including goal setting as part of health-, lifestyle- or treatment-related interventions;
2. Studies focusing uniquely on students and the education context.



## Step 4. Search strategy: How should the studies be sought?

Based on the question, you next have to conduct a structured search for all relevant studies in the international research literature. In the first instance, you should concentrate your search on relevant bibliographical databases using clearly defined search terms. At the very least, conduct your search using ABI/INFORM from ProQuest and Business Source Premier from EBSCO. Depending on the RR question, you may also need to search in databases that are aimed at neighbouring disciplines such as psychology (PsycINFO), education (ERIC) or healthcare (PubMed).

### EXAMPLE

The following four databases were used to identify studies: ABI/INFORM Global, Business Source Premier, PsycINFO and Web of Science. The following generic search filters were applied to all databases during the search:

1. Scholarly journals, peer-reviewed
2. Published in the period 1980 to 2016 for meta-analyses and the period 2000 to 2016 for primary studies
3. Articles in English

A search was conducted using combinations of different search terms, such as ‘goal setting’, ‘goal attainment’, ‘goal pursuit’ and ‘performance’. In addition, the references listed in the studies retrieved were screened in order to identify additional articles for possible inclusion in the REA. We conducted 8 different search queries and screened the titles and abstracts of more than 350 studies. An overview of all search terms and queries is provided in Annex I.

### Unpublished studies

In an REA, the search is often limited to studies published in peer-reviewed journals. However, to reduce publication bias you should also search for study findings reported in:

- conference papers and proceedings;
- dissertations and theses; and
- working papers.

When relevant studies are found, you should take a random ‘dip’ sample to determine whether the findings reported are notably different from those documented in studies published in peer-reviewed journals. If this is the case, this should be considered as a strong indication of publication bias, and it should be reported in your REA.

## Documentation of the search

It is important to carry out a search that is transparent, verifiable and reproducible. For this reason, you should clearly document the search process, preferably in the form of a table that shows the search terms used, how search terms were combined, and how many studies were found at every step. In addition, the table should specify the date on which the search was conducted and the search filters that were applied. An example is provided below.

ABI/Inform Global, Business Source Premier, PsycINFO, peer reviewed, scholarly journals, July 2014			
Search terms	ABI	BSP	PSY
<b>S1:</b> ti("virtual team*") OR ti("virtual work*") OR ab("virtual team*") OR ab("virtual work*")	57	38	59
<b>S2:</b> ti("telework**") OR ab("telework**") OR ti(telecommut* OR ab(telecommut*)	98	87	120
<b>S3:</b> ti("mobile team**") OR ti("mobile work**") OR ab("mobile team**") OR ab("mobile work**")	52	48	62
<b>S4:</b> ti("remote team**") OR ti("remote work**") OR ab("remote team**") OR ab("remote work**")	12	52	14
<b>S5:</b> ti("distributed team**") OR ti("distributed work**") OR ab("distributed team**") OR ab("distributed work**")	42	34	73
<b>S6:</b> S1 – S5	153	167	210
<b>S7:</b> S6 AND ab(experiment* OR laboratory OR "field study" controlled OR "control group" OR "comparison group" OR "control variable" OR quasi OR longitudinal OR randomized	68	34	42
<b>Duplicates removed</b>	48		

## Step 5. Study selection: How should you select the studies?

In general, a search will yield a large number of studies – sometimes several thousand. Some studies will not be directly relevant to the research question and PICOC. Hence, the next step is to screen them to check whether they meet the inclusion criteria. In an REA, studies are screened by only one reviewer. Therefore, to ensure the trustworthiness of the REA findings, a second reviewer should take a random sample from the studies included and independently review them for inclusion criteria and practical relevance. To quantify the degree of agreement with the first reviewer, the inter-rater reliability (IRR) should be assessed (Hallgren, 2012). When the IRR's magnitude indicates that the extent of agreement among the reviewers is poor, the second reviewer should independently screen the total sample of studies included. Any disagreement should be solved by discussion and consensus, or by consulting a third reviewer.

Screening is usually a two-stage process, the first involves reviewing the abstracts and the second reviewing the full studies.

### Review abstracts

You should read the abstracts that you have found through your searches. Compare each abstract against the inclusion criteria, and, if the abstract meets the criteria, then you should read the full study. Not all abstracts will contain information on all the inclusion criteria (this is particularly a problem with electronic searching). In these cases, you need to make decisions as to whether or not to include the study based on the limited information available. In cases of doubt, lack of information, or disagreement between the two reviewers, the study should be included.

### Review full studies

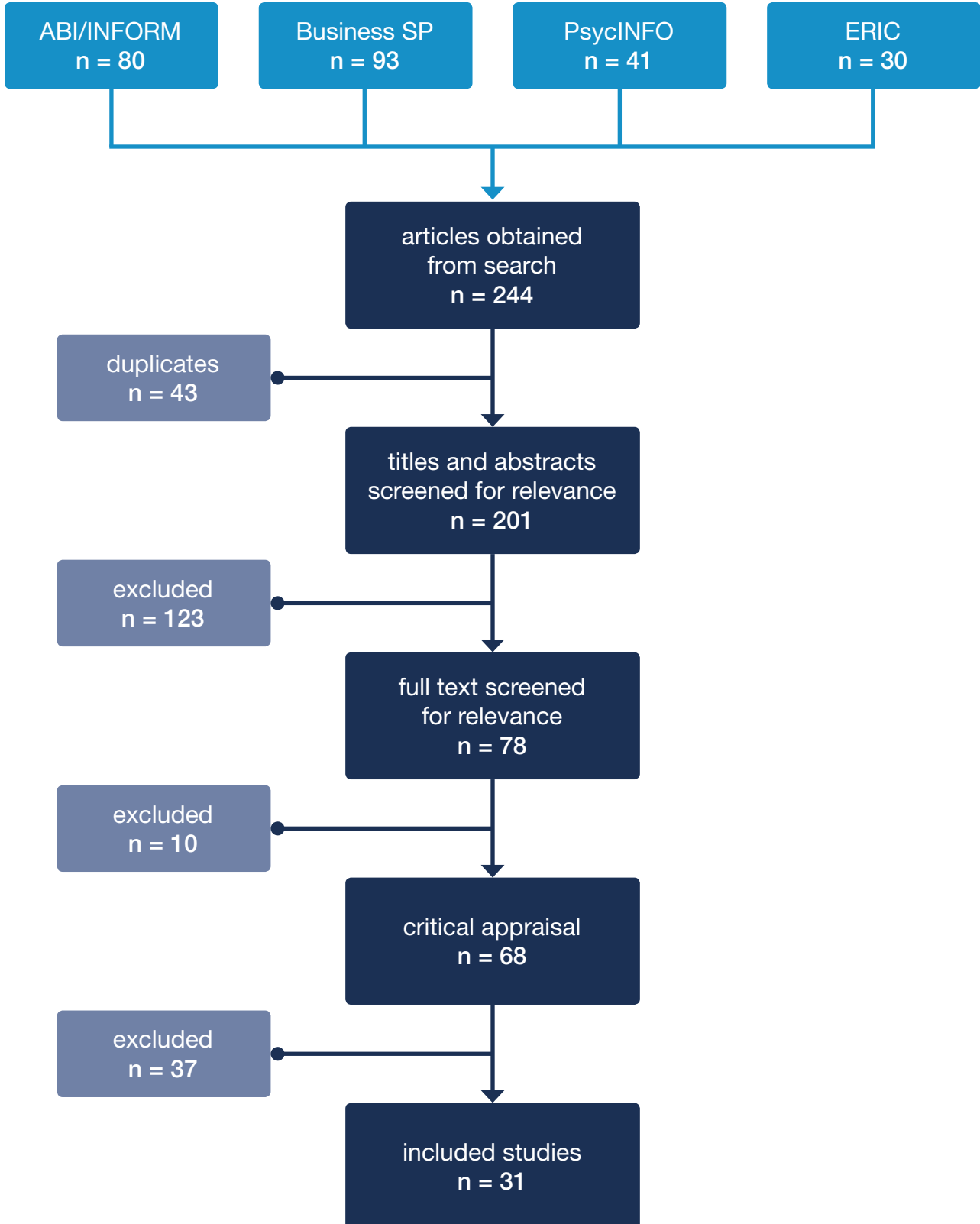
You should read the full article and compare it against the inclusion criteria. Where two reviewers disagree on inclusion, a third reviewer should assess the study's appropriateness for inclusion with no prior knowledge of the initial reviewers' assessments. The decision of the third reviewer should be final.

#### EXAMPLE

Selection took place in two phases. First, the titles and abstracts of 350+ studies were screened for their relevance to this review. In case of doubt, the study was included. Duplicate publications were removed. This first phase yielded 50 meta-analyses and 30 controlled and/or longitudinal studies.

## Documentation

You must clearly document the selection process. You could, for example, use a flowchart to show how many studies are excluded, as in the example below.



## Step 6. **Data extraction: What information should you extract?**

Data extraction involves the collation of the results and other information of the studies included. From each study, you should report, preferably in the form of a clearly structured table, any information relevant to the REA question, such as year of publication, research design, sample size, population (e.g., industry, type of employees), outcome measures, possible moderators or mediators, main findings, effect sizes with a 95% confidence interval, limitations, and the final level of trustworthiness (see step 7). An example is provided on the next page.



Author & year	Sector / Population	Design + sample size	Main findings	Effect size	Limitations	Level
Abraham & Graham-Rowe (2009)	Systematic review; 2/3 RCT; 1/3 quasi-experimental	Worksites; 8 studies; N = 624 employees	Worksite physical activity interventions which include specific goal setting, goal reviews (i.e. follow-up) and graded tasks have a small, positive impact on fitness-related outcomes	small, positive, $d = .33$ if specific goal-setting; $d = .37$ if inclusion of goal review; $d = .44$ if inclusion of graded tasks	Limited relevance to the review question	AA
Bandura & Locke (2003)	General population	Traditional literature review	Discusses the importance of self-efficacy for understanding, predicting and changing people's performance or goal attainment. Self-efficacy is stated to be related (based on meta-analytical findings from previous studies), among others, to more proactive (self-set) goal-setting, challenging goals, and faster goal attainment, as well as effort and performance.	No effect sizes provided	No systematic search, no information regarding design of included studies	D
Brown, 2005	Canadian government employees in a training program	Randomized controlled trial, N=74, field setting	Both participants who were urged to do their best and those who set proximal (shorter-term) as well as distal (= longer-term) goals had increased transfer of training (= maintenance of learned material over time and generalization of learned material from the classroom to the workplace context) relative to those who set only distal outcome goals. There was no significant difference in the transfer level of participants urged to do their best and those who set proximal plus distal goals. In addition, there was no difference between the experimental conditions regarding the effect on self-efficacy. This suggests that support the conclusion that distal outcome goals are not effective in bringing about an increase in transfer when participants are learning new skills	$\text{Eta}^2 = 0.11$	Short time frame between training and measurement (six weeks)	A
Fu, 2009	Industrial sales people	Before after, with double post-test (3 months and 6 months) N=143	The study indicates further that self-set goals fully mediate the relationship between assigned goals and selling effort (ass goals impact ssg and then selling effort). In addition, the longitudinal data indicate that company-assigned goals, self-set goals, and selling effort all positively influence future new product sales, but not self-efficacy (not significant). Interestingly, the results of the study fail to confirm an inverted, U-shaped relationship between assigned goals and effort	Impact on new product sales: assigned goals ( $\beta = 0.24$ ), SSG ( $\beta = 0.34$ ), and effort ( $\beta = 0.20$ )	No serious limitations	C
Schweitzer et al, 2004	Undergraduate studies, US	RCT, n=159	Results of a laboratory experiment utilizing high, low, increasing, decreasing, and "do your best" goal structures across multiple rounds provide evidence that depletion mediates the relationship between goal structures and unethical behavior, and that this effect is moderated by the number of consecutive goals assigned.	Very small effect sizes	Artificial students	A



## Step 7. Critical appraisal: How to judge the quality of the studies.

### Methodological appropriateness

You can usually find a study to support or refute almost any theory or claim. It is thus important that you determine which studies are trustworthy (i.e. valid and reliable) and which are not. You should first determine the trustworthiness of a scientific study by its methodological appropriateness. For cause-and-effect claims (i.e. if we do A, will it result in B?), a study has a high methodological appropriateness when it fulfills the three conditions required for causal inference: co-variation, time-order relationship, and elimination of plausible alternative causes (Shaughnessy & Zechmeister, 2006). A study that uses a control group, random assignment and a before-and-after measurement is therefore regarded as the ‘gold standard’ for effect studies<sup>1</sup>.

Non-randomized studies and before-after studies come next in terms of appropriateness. Cross-sectional studies (surveys) and case studies are regarded as having the greatest chance of showing bias in their results and thus fall lower in the ranking of appropriateness. Systematic reviews or meta-analyses based on pooled results of randomized controlled studies are thus regarded as the most appropriate design.

To determine the methodological appropriateness of effect studies and impact evaluations, we recommend using the following six levels of appropriateness, which are based on the classification system of Shadish, Cook and Campbell (2002), and Petticrew and Roberts (2006).

Design	Level
Systematic review or meta-analysis of randomized controlled studies	<b>AA</b>
Systematic review or meta-analysis of non-randomized controlled and/or before-after studies	<b>A</b>
Randomized controlled study	
Systematic review or meta-analysis of controlled studies without a pretest or uncontrolled study with a pretest	<b>B</b>
Non-randomized controlled before-after study	
Interrupted time series	
Systematic review or meta-analysis of cross-sectional studies	<b>C</b>
Controlled study without a pretest or uncontrolled study with a pretest	
Cross-sectional study (survey)	<b>D</b>
Case studies, case reports, traditional literature reviews, theoretical papers	<b>E</b>

<sup>1</sup> It should be noted that randomized controlled studies are often conducted in an artificial (lab-type) setting – with students carrying out prescribed work tasks – which may restrict their generalizability. Non-randomized studies in a field setting – with employees carrying out their normal tasks within an organizational setting – on the other hand, have a lower level of trustworthiness, but can still be useful for management practice.

Note, however, that the level of methodological appropriateness as explained above is only relevant in assessing the validity of a study that examines a cause-and-effect relationship that might exist between an intervention (e.g. performance appraisal) and its outcomes (e.g performance). When the REA question concerns a non-effect or non-impact question, for example the prevalence of phenomenon ('How many / often do people / organizations ...?'), a cross-sectional study may be the most appropriate design (Petticrew and Roberts, 2003).

## Methodological Quality

A study's trustworthiness, however, is also affected by its methodological quality, that is, the way it was conducted. To determine methodological quality, you should systematically assess all the studies included based on explicit quality criteria, such as the PRISMA (Moher, Liberati, Tetzlaff, & Altman, 2009) and CONSORT statement (Moher, Schulz, & Altman, 2001), the CASP checklists (Critical Appraisal Skills Programme, n.d.), or the checklists of the EPPI-Centre (Newman & Elbourne, 2005). For management and organizational studies, you can use CEBMA's CAT Manager app, which you can download from Google Play (Android) or App Store (iPhone).

Based on a tally of the number of weaknesses, you might downgrade the level of trustworthiness by one or more levels. Examples of possible weaknesses are small sample size or low scale reliability. To determine the final level of trustworthiness use the following rule of thumb:

- **1 weakness** = no downgrade (we accept that nothing is perfect)
- **2 weaknesses** = downgrade 1 level
- **3 weaknesses** = downgrade 2 levels
- etc.

In an REA, the methodological appropriateness and quality of the studies included are appraised by only one reviewer. Therefore, to ensure the trustworthiness of the REA findings, a second reviewer should take a random 'dip' sample from the studies included and independently review each study's methodological appropriateness and quality. To quantify the degree of agreement with the first reviewer, the inter-rater reliability (IRR) should be assessed (Hallgren, 2012). As explained in Step 5, when the IRR's magnitude indicates that the extent of agreement among the two reviewers is poor, the second reviewer should independently critically appraise the total sample of studies included. Again, any disagreement should be solved by discussion and consensus, or by consulting a third reviewer.

## EXAMPLE 1

After critical appraisal of 24 studies, only four studies were included. Most studies were excluded due to serious methodological shortcomings. One of the studies included was a systematic review, representing the results of 18 studies. The overall quality of the included studies, however, was low. For instance, all but two of the studies included in the systematic review were self-report surveys, and due to heterogeneity across studies it was not possible to calculate a pooled estimate of effect. The three primary studies used a cross-sectional design. As a result, the trustworthiness of the scientific evidence supporting the following main findings is limited.

## EXAMPLE 2

The overall quality of the studies included was high. Of the 34 meta-analyses, 13 included randomized and/or non-randomized controlled studies and were therefore qualified as level A or AA. Of the remaining 21 meta-analyses 1 was graded as level B, and 14 were graded as level C. Finally, 6 studies were graded level D because they were traditional literature reviews. It should be noted, however, that some meta-analyses were graded level C because it was insufficiently clear what type of studies were included. The actual level of evidence of these meta-analyses (and the overall quality of the studies included in this REA) may thus be higher.

Most of the secondary studies were based on cross-sectional studies and thus graded level B or lower. Only seven qualified as level A. Of the 19 primary studies, 15 qualified as randomized controlled studies and thus graded level A. The remaining 4 studies concerned quasi-experimental or longitudinal designs and were graded level B or lower.

## Effect Sizes

An outcome can be statistically significant, but it may not necessarily be practically relevant. Note that even a trivial effect can be statistically significant if the sample size is large. This works the other way around as well: even a large, practical relevant effect can be statistically non-significant if the sample size is small. Also, keep in mind that p-values do NOT measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone (Ziliak, 2016). Nor can a p-value tell you the size of an effect, the strength of the evidence or the importance of a result.

For this reason, you should pay little attention to the p-value but instead assess the ‘effect size’ – a standard measure of the magnitude of the effect – of the studies included when addressing impact questions.

To determine the magnitude of an effect, apply Cohen’s rules of thumb (Cohen, 1988; see below). According to Cohen a ‘small’ effect is an effect that is only visible through careful examination. A ‘medium’ effect, however, is one that is ‘visible to the naked eye of the careful observer’. Finally, a ‘large’ effect is one that anybody can easily see because it is substantial.

Effect size	Small	Medium	Large
Standardized mean difference: $d, \Delta, g$	$\leq .20$	.50	$\geq .80$
Correlation: $r, \rho$	$\leq .10$	.30	$\geq .50$
Correlation: $r^2$	$\leq .01$	.09	$\geq .25$
ANOVA: $\eta^2, \omega^2$	$\leq .01$	.06	$\geq .14$
Chi-square: $\omega^2$	$\leq .10$	.30	$\geq .50$
Simple regression: $\beta$	$\leq .10$	.30	$\geq .50$
Multiple regression: $\beta$	$\leq .20$	.50	$\geq .80$
Multiple regression: $R^2$	$\leq .02$	.13	$\geq .26$

Note, however, that Cohen’s rules of thumb were meant to be exactly that – ‘rules of thumb’ – and are for many reasons arbitrary (Cooper & Lindsay, 1998). For example, a standard mean difference of .20 may be regarded as ‘small’ when the outcome concerns job satisfaction but ‘large’ when the outcome concerns fatal medical errors. When assessing impact, it is therefore important to relate the effect size directly to the outcome measured.



## Step 8. Results: What did you find?

### Step 8.1. Definition: What is meant by X?

Most REA questions include one or more key elements / constructs, for which several definitions are available. In this step, you should provide an overview of the most common definition(s).

#### EXAMPLE 1

An incentive is commonly defined as “something that arouses action or activity” (<http://www.merriam-webster.com>). In the domain of management, incentives can be defined as “...plans that have predetermined criteria and standards, as well as understood policies for determining and allocating rewards” (Greene, 2011, p. 219). Incentives include all forms of rewards (and punishments) that are based on an employee’s performance or behavior. Promotions, grades, awards, praise, and recognition are therefore all incentives. However, financial incentives such as money, bonus plans, or stock options are the most commonly used (Cerasoli, Nicklin, & Ford, 2014). Formally, incentives differ from rewards. Incentives refer to all stimuli that are provided in advance, whereas rewards are offered after a given performance (Garbers & Konradt, 2014). In the scientific literature and management practice, however, these terms are used interchangeably.

#### EXAMPLE 2

The concept of self-managing teams is referred to in various ways, using terms such as ‘autonomous groups’, ‘shared’ or ‘self-directed teams’; all of these terms refer to teams that are hallmarked by autonomy. We use the term ‘self-managing teams’ to cover all of the different descriptions of this concept. Ingvaldsen & Rolfsen (2012) refer to the standard definition of autonomous groups as “groups responsible for a complete product or service, or a major part of a production process. They control members’ task behavior and make decisions about task assignment and work methods” (Cummings and Worley, 2005: p. 341). What’s more, self-managing teams are seen as an option in organizational design, including teams with broad authority to decide how to execute certain tasks (Johnson et al., 2013).

## Step 8.2. Causal mechanism: How is X assumed to have an effect on Y?

A causal mechanism spells out the process by which an intervention, construct or phenomenon is expected to have an effect on a certain outcome. The causal mechanism is often based on one or more (social or psychological) theories that explain why the assumed effect occurs and under which circumstances.

### EXAMPLE 1

The assumed causal mechanism of performance appraisal is based on two theories: social comparison theory (Festinger, 1954), and feedback intervention theory (Kluger & DeNisi, 1996). Social comparison theory suggests that individuals tend to compare themselves with others in order to make judgments regarding their performance. They are concerned not only about their performance in an absolute sense, but also about how they measure up in relation to relevant peers. In addition, this theory posits that individuals have a strong desire to improve their performance when faced with unfavorable comparative information. Feedback intervention theory suggests that when confronted with a discrepancy between what they wish to achieve and the feedback received, individuals are strongly motivated to attain a higher level of performance. The practice of performance appraisal therefore assumes that informing an employee about the discrepancies between the organization's standard and his/her current performance – implying that he/she is achieving lower than most other colleagues – will motivate the employee to achieve a higher level of performance.

### EXAMPLE 2

The construct of perceived supervisory support stems from the norm of reciprocity, which states that people treat others as they would like to be treated, repaying kindness with kindness and retaliating against those who inflict harm (Brunell et al., 2013; Gouldner, 1960). Put differently, when a manager helps his or her employees in times of need or recognizes their extra efforts, employees will feel inclined to act to benefit the manager (e.g., by meeting goals and objectives) and thus the organization as a whole (Edmondson & Boyer, 2013; Eisenberger, Huntington, Hutchison, & Sowa, 1986). Not surprisingly, physicians experiencing inadequate supervisory support tend to describe their commitment to the hospital and patients in negative terms, which negatively affects their performance (Tayfur & Arslan, 2013).

## Step 8.3. Main findings

In this section, you should provide an overview of the main findings relevant to the REA question. For each finding, you should present the main evidence from the REA, including its level of trustworthiness and effect size.

### EXAMPLE 1

*Goal setting has a moderate positive effect on performance (level AA)*

High-quality meta-analyses (level AA) in a wide range of disciplines (management, medicine, sports, rehabilitation, prevention, etc.) and populations (patients, athletes, managers, senior adults, children, etc.) have demonstrated the positive effects of goal-setting interventions on performance outcome. Adding to the robustness of this finding is the fact that goal-setting, as an intervention, also affects behavior or achievement in areas other than workplace performance. Meta-analyses (level A) suggest that, among others, goal-setting interventions can increase learning (Mesmer-Magnus & Viswesvaran, 2007; Sitzmann & Ely, 2011), job search success (Liu, Huang & Wang, 2014), training transfer (Rahyuda, Syed & Soltani, 2014), well-being (Klug & Maier, 2015), physical activity (McEwan et al., 2016; O'Brien et al., (2015) and fitness-related outcomes (Abraham & Graham-Rowe, 2009).

### EXAMPLE 2

*Perceived supervisory support has a moderate positive effect on performance and a strong positive effect on job satisfaction and organizational commitment (level A)*

Perceived supervisory support was found to have a moderate positive effect on performance and a strong positive effect on organizational commitment (Edmondson & Boyer, 2013). Because the results were consistent throughout all the meta-analyses, evidence level A was assigned. In addition, all meta-analyses that studied the relationship between perceived supervisory support and job satisfaction found positive effects (Edmondson & Boyer, 2013; Luchman, & Gonzáles-Morales, 2013; Mor Barak et al., 2009; Ng & Sorensen, 2008). Several cross-sectional studies, albeit with lower levels of evidence, confirmed this finding (Elias & Mittal, 2011; Mansell, Brough & Cole, 2006).

### EXAMPLE 3

*The use of a guiding coalition is widely advocated in the popular literature, but the assumed effect on successful change is only partially supported by the scientific literature (level D)*

Although the use of a guiding coalition in organizational change is widely advocated, we found only one study – a cross-sectional survey – that showed a moderate correlation between the success of the organizational change and the existence of a powerful guiding coalition that directed the change (Morris, Griffin & Crawford, 1999, pp. 742-743). The qualitative case studies stressed the importance of the formation of a guiding coalition as a starting principle for successfully achieving change. However, the effect on the outcome was not measured. (McCracken & McIvor, 2012; Loeser, O’Sullivan & Irby, 2007; Appelbaum, Habashy, Malo & Shafiq, 2012).

### EXAMPLE 4

*Overall, financial incentives have a moderate positive effect on performance (level A)*

There is strong evidence that financial incentives tend to have a moderate positive effect on performance (Cerasoli, et al., 2014; Garbers & Konradt, 2014; Weibel, et al., 2010). This positive effect is often referred to as the ‘price effect’: the financial incentive increases the intention to perform well because of the monetary benefit. However, this effect differs between forms of incentives, types of motivation, and performance outcomes. Financial incentives can also impair performance when set so high that they interfere with learning goals and cognitive processes required for task performance (Arieley et al., 2009).



## Step 8.4. Moderators and mediators

The relationship between a certain intervention, construct or phenomenon and a certain outcome is often complicated and is contingent upon a large number of moderators and mediators. A moderator is a variable that affects the direction and/or strength of the relation between an independent or predictor variable (e.g. goal setting) and an outcome variable (e.g. work performance). Put differently, moderators indicate when or under what conditions a particular effect can be expected. For this reason, they are also referred to as ‘boundary conditions’. A mediator is a variable that specifies how or why a particular effect or relationship occurs. Thus, if you remove the effect of the mediator, the relationship between the independent or predictor variable (in this case performance appraisal) and the outcome variable (work performance) will no longer exist. In short, moderators specify when a certain effect will hold, whereas mediators determine how or why the effect occurs. As a result, the key question is not ‘What is the effect of X on Y?’, but ‘Given the target group, the objectives and the context involved, what are the factors moderating or mediating the effect of X that need to be taken into account?’ In this section, we present the main moderators and mediators, including their level of trustworthiness and effect size.

### EXAMPLE 1

*The effect of goal-setting varies across workers’ ability levels (level A)*

A recent randomized controlled study found that low-ability workers for whom goals were likely to be challenging increased their performance by 40% in the goal-setting treatment with respect to the baseline while high-ability workers achieved the same level of performance across treatments (Corgnet, 2012). This finding confirms the outcome of a previous randomized controlled study that ‘ability-based’ goals are more effective at improving performance than a ‘one-size-fits-all’ approach, where everyone is assigned the same performance target (Jeffrey, 2012).

### EXAMPLE 2

*Raising fear leads employees to engage in short-term thinking (level A)*

Fear can be aroused in people with a specific goal in mind; to take preventive measures, or to stimulate proactive behavior. However, there is strong evidence indicating that fear (e.g. job insecurity) tends to lead people to engage in short-term thinking, favoring immediate consequences (Gray, 1999).



### EXAMPLE 3

*The outcome of managers' own performance appraisal has a large effect on the way in which they evaluate their employees (level A).*

AA combination of studies (including a randomized controlled study) demonstrated that managers who receive positive feedback about their performance subsequently rate their employees significantly higher than do managers who receive negative feedback regarding their own performance (Latham et al., 2008). Surprisingly, this effect even occurred when managers knew their own evaluation was bogus.

### EXAMPLE 4

*An unfair process has a small to medium negative effect on employee behavior further to a breach of the psychological contract (level D)*

A cross-sectional study (Kickul, Lester & Finkl, 2002) found that procedural justice was an important determinant of behavior following breaches of external outcomes (e.g. pay) in terms of the psychological contract (beliefs regarding mutual obligations) between the employer and the employee. When procedural justice was perceived to be low, employees showed less organizational change behavior, satisfaction, as well as in-role job performance.

## Step 9. Synthesis: What does it all mean?

In the synthesis you should combine all separate findings into a coherent summary of the evidence that answers the REA question and enables conclusions to be drawn. The methods that you should employ for the synthesis depend on the nature and comprehensiveness of the evidence. At the very least you should describe two aspects in your synthesis:

- The volume and characteristics (e.g. comprehensiveness and methodological quality) of the overall evidence base
- What the evidence indicates regarding the REA question(s)

### Describing the volume and characteristics of the overall evidence base

Providing a description of the volume and characteristics of the evidence found by the REA will enable you to determine the adequacy of the overall evidence base to answer the REA question. The details should be specific to each question and could include:

- Research designs used
- Variables studied
- Outcomes measured
- Details of context (for example population, geographical region, industry, time period)

In addition, you should make statements regarding the consistency and convergence of the evidence. Examples could include:

- **Consistent evidence** = A range of different forms of evidence point to identical or similar conclusions
- **Contested evidence** = One or more study/studies directly refutes or contest the findings of other studies, raising questions about the trustworthiness of the purported effect.
- **Mixed evidence** = Studies based on a variety of different designs or methods, applied in a range of contexts, have produced results that suggest underlying differences in the nature of the effects observed or important differences across studies that are not yet well-understood. (N.B. Often associated with unmeasured heterogeneity in studies, treatments and populations.)

### EXAMPLE 1

Over the last 50 years, there has been a considerable number of studies on the topic of performance appraisal. In fact, performance appraisal may well be one of the most widely studied topics in the domain of management, with research dating back at least to the early 1920s and continuing to the present. After critically selecting and assessing the available empirical studies, we can conclude that the scientific evidence is rich in both quantity and quality. In addition, the assumed positive effect of performance appraisal on work performance is grounded in three well-established social psychological theories. However, this REA did not yield any randomized controlled studies that measured the direct effect of performance appraisal on workplace performance. Instead, the best available evidence consists of a large number of high-quality studies that focus on (one or multiple) separate elements of the appraisal process, such as rating, judgment, feedback or perceived fairness. The outcome of these studies suggests that the relationship between performance appraisal and work performance is contingent upon a large number of moderators and mediators.

### EXAMPLE 2

Over the past decades there have been numerous publications reporting the high failure rates of change initiatives. None of them, however, provides any valid, reliable empirical evidence supporting this claim.

### EXAMPLE 3

Over the last 50 years, there have been a considerable number of studies on the relationship between transactional leadership and (change) performance. After critically selecting and assessing the available scientific evidence, the main findings are mainly based on well-designed meta-analyses and randomized controlled studies.

### EXAMPLE 4

Resistance to change is considered to be one of the most important topics of change management, as shown by the wide coverage of this subject in both the popular management literature as well as in academic publications. Many people have made claims about resistance to change, which is seen as prominent factor contributing to the failure of organizational change initiatives. The results of this rapid evidence assessment (REA) however, leave us with a sense of disappointment. There are hardly any well conducted, high-quality studies and all of the studies included in this REA suffer from serious methodological shortcomings.

## Describing what the evidence indicates

In the synthesis you need to consider what the evidence indicates in relation to the REA question. Again, how to do this will be context-specific, depending on the question and evidence base. However, for an REA it is likely that you would give a narrative synthesis, as opposed to a quantitative synthesis or meta-analysis, which tend to be beyond the scope of most REAs. Narrative syntheses primarily rely on the use of words and text to summarize findings from multiple studies. However, you can also use tables and graphical descriptions to support narrative descriptions.

### EXAMPLE 1

... The outcome of this REA indicates that employee reaction to feedback is one of the most important mediators in the performance-appraisal process. In fact, there is strong evidence that employees' reaction to feedback, and not feedback per se, determines the extent to which their performance will improve. How an employee will react to the feedback on his/her appraisal, however, is strongly moderated by the perceived fairness of the appraisal process. When the procedure is perceived to be just, employee reactions are more likely to be favorable, largely irrespective of their performance ratings. Perceived fairness, in turn, is moderated by several other variables, of which perceived usefulness, rating method, rating accuracy, focus of the feedback, level of employee participation, and quality of the relationship between manager and employee seem to have the largest impact. A second factor that moderates employees' reaction to the feedback embraces personality variables such as self-esteem, locus of control and openness to feedback...

### EXAMPLE 2

... After critical quality assessment of the scientific evidence, we can safely conclude that financial incentives have a positive impact on performance, also known as the 'price effect'. However, financial incentives also have a negative impact on the intrinsic motivation of employees, which is known as the 'crowding-out' effect. The net result of these two opposing effects determines a possible gain or loss in performance. In addition, the net effect is influenced by several mediating and moderating variables. For instance, intrinsic motivation, often linked to executing interesting tasks, is critical when quality outcomes determine performance. In addition, intrinsic motivation is likely to be reduced by the use of direct financial incentives. Direct financial incentives, however, tend to have a positive effect on employees' quantity performance, associated with non-interesting tasks. Finally, when entire teams are incentivized, the evidence suggests that the rewards should be distributed equitably among team members, therefore strengthening employees in their motivation and diminishing the risk of freeriding...

### EXAMPLE 3

... Overall there is strong evidence demonstrating that commitment is positively correlated to several organizational outcomes, such as job performance. When affective and normative commitment are high, performance is likely to increase, especially contextual job performance. A remarkable finding is that the relationship between affective commitment and performance depends on an employee's tenure, and that continuous commitment seems to have a negative effect. However, the effect sizes are small to moderate, and there are some indications that organizational trust might be a better predictor...

### EXAMPLE 4

... Although goal-setting can be regarded as one of the most effective organizational interventions, this REA also demonstrates that goal-setting should not be used as a one-size-fits-all, over-the-counter treatment to boost performance, as there are several moderators that affect the outcome. For example, when employees must first acquire knowledge or skills to perform a task, or when the task involved is complex, then specific and challenging goals can have a negative effect on performance. In those situations, behavioral goals and learning goals are more effective, as outcome goals only result in increased performance once people have mastered the task. Furthermore, this REA indicates that the effect of goal-setting varies across workers' ability levels, implying that 'ability-based' goals will be more effective than a 'one-goal-for-all' approach, where everyone is assigned the same performance target...

## Step 10. Conclusion

You should make the conclusion of your REA a concise statement (of two or three sentences) on the main findings on the REA question.

### EXAMPLE 1

The scientific literature does not support the claim that organizational change requires leaders with strong emotional intelligence.

### EXAMPLE 2

Goal-setting is one of the most powerful and evidence-based interventions for enhancing performance, provided that moderating factors such as goal attribute, type of task, organizational context and employee characteristics are carefully taken into account.

### EXAMPLE 3

The scientific research literature supports my assumption that a fair change process is important to realizing successful change, given the moderate positive effect of procedural justice on organizational outcomes. Although the effects are mostly small to medium, the indications are that there is indeed a positive relationship between procedural justice and acceptance, commitment and behavior in the specific context of organizational change.



## Step 11. Limitations

In an REA you are aiming to provide a balanced assessment of what is known (and not known) in the scientific literature about an intervention or practical issue by using a systematic methodology to search and critically appraise empirical studies. Nevertheless, all REAs have limitations. In your REA you should explicitly describe any limitations and discuss how they possibly impacted the findings of the assessment. Below is an example of a description of limitations that are inherent to REAs.

### EXAMPLE

To provide a 'rapid' review, concessions were made in the breadth and depth of the search process. As a consequence, some relevant studies may have been missed.

A second limitation concerns the critical appraisal of the studies included: this REA did not incorporate a comprehensive review of the psychometric properties of the tests, scales and questionnaires used.

A third limitation concerns the fact that the evidence on several moderators is often based on a limited number (sometimes only one) of studies. Although most of these studies were well controlled or even randomized, no single study can be considered to be strong evidence – it is merely indicative.

Finally, this REA focused only on high-quality studies, i.e. studies with a control group and/or a before- and after-measurement. For this reason, usually a large number cross-sectional studies are excluded. As a consequence, new, promising findings that are relevant for practice may have been missed.

Given these limitations, care must be taken not to present the findings presented in this REA as conclusive.



## Step 12. Implications of the findings for practice

Once you have used the evidence found to answer the REA's main question, you should use the final part of the assessment to relate the findings to the background of the REA and the PICOC described in step 1 and 2. For example: Is the evidence supportive of current practice? What are the estimated benefits and harms? What, based on the evidence found, are your specific recommendations for action? Importantly, how might you consider the concerns and interests of different organizational stakeholders in your recommendations?

### EXAMPLE 1

The fair process effect in organizations is observed when change leaders increase the degree to which the decision-making process is consistent, accurate, unbiased, and open to employee voice and input. When procedural justice is not taken into account, employees may feel treated unfairly and resistance may increase. To actively design a fair change process, the six classic criteria for procedural justice of Leventhal (1980) – still used by many researchers – may serve as a useful checklist. These criteria can be turned into practical guidelines for the purpose of organizational change as follows: (a) the change approach needs to be consistently applied to all employees at all times; (b) it needs to be impartial, i.e. prejudice or stereotyping are eliminated; (c) the information on which decisions are based needs to be accurate; (d) if the situation demands it, opportunities should be provided to correct or change plans or processes; (e) those responsible for the organizational change (the change managers/executives responsible) need to represent the interests of all stakeholders affected by the change; and (f) the ethical standards and values of those involved should never be disregarded.

## EXAMPLE 2

Financial incentives can be used to increase the employee motivation and performance needed in changing circumstances. However, upper management should have a clear vision about the change in performance or behavior that it desires, as it requires different approaches to incentivizing. Intrinsically motivated employees executing interesting tasks where quality outcomes usually determine success could be encouraged by indirect incentives. Direct financial incentives are effective when extrinsic motivation and quantitative performance need to be promoted. The Executive Board should therefore continuously ‘calculate’ the proposed net effect (positive price effect versus negative crowding-out effect) when defining a pay plan. Ultimately, if the plan is designed to increase team performance, not all incentives should be distributed equally, as this may harm individual motivation. Some attention should be given to rewarding individual contributions as well. At the same time, an overemphasis on financial incentives in place of other rewards related to intrinsic motivation and personal development can lead to taking of inappropriate risks in pursuit of financial gain. Given the risk of unintended consequences from financial incentives it is important to assess periodically the effects such incentives have external stakeholders such as customers and patients as well as internal stakeholders such as employees in customer-facing and indirect-service roles.

## EXAMPLE 3

Since the research evidence does not provide us with a decisive answer to the question whether self-managing teams perform better, no clear recommendations for practice can be given on whether our organization should implement self-managing teams or not. Neither can it be indicated whether self-managing teams will help our organization to change more successfully or not. In this light, we suggest that our organization needs to be extra careful about implementing self-managing teams when the divisions involved have high levels of hierarchy (e.g. micro-level centralization) or organizational formalization. Divisions typically defined by high-tech novelty and radical innovation, however, do not need to be as cautious in implementing self-managing teams, although they do need to be aware of the potential for conflict within their teams and the negative impact that this could have on team performance. The available knowledge on these contextual factors is still limited due to the current embryonic state of research evidence on this subject. Any introduction of self-managing teams that may be planned should therefore be considered carefully.

#### EXAMPLE 4

This REA demonstrates that Emotional Intelligence (EI) is not a radical new construct that will redefine leadership. Even though EI has (some) positive effects, these effects can also be explained by the overlap with other psychological constructs. In addition, the claims made by well-known consultancy firms such as Hay Group that “EI can make the difference between a highly effective and an average professional contributor” is not supported by the outcome of this REA. For this reason, we advise against investing in training courses that claim to develop our executives’ EI.



# Checklist

1. Have you clearly described the background and context of the REA question	
2. Does the REA address a clearly focused question? Is it clear what the REA will answer?	
3. Have you clearly defined the inclusion criteria (e.g. population, outcomes of interest, study design)?	
4. Have you conducted a comprehensive literature search using relevant research databases (i.e. ABI/INFORM, Business Source Premier, PsycINFO, Web of Science, etc.)?	
5. Is the search systematic and reproducible (e.g. were searched information sources listed, were search terms provided, were search results reported)?	
6. Have you searched for study findings reported in unpublished or 'grey' literature to determine the presence of publication bias?	
7. Have you selected the studies using explicit inclusion and exclusion criteria and documented the selection process (e.g. using a flowchart to show how many studies you have excluded and why)?	
8. Did a second reviewer take a random sample from the studies included and independently review them for inclusion criteria and practical relevance? Was the inter-rater reliability assessed?	
9. Have you clearly described the key features (year of publication, population, sample size, study design, outcome measures, effect sizes, limitations, level of trustworthiness) of all studies included?	
10. Have you assessed the methodological appropriateness and methodological quality of each study using predetermined quality criteria?	
11. Did a second reviewer take a random sample from the studies included and independently review their methodological appropriateness and quality? Was the inter-rater reliability assessed?	
12. Have you provided definitions of the key elements/constructs in the REA question?	
13. Have you clearly described the assumed causal mechanism?	
14. Have you provided an overview with the main findings, including their level of trustworthiness and effect size?	
15. Have you provided an overview with the main moderators and mediators, including their level of trustworthiness and effect size?	
16. Does the synthesis describe the volume and characteristics of the overall evidence base?	
17. Does the synthesis combine the separate findings into a coherent summary of the evidence?	
18. Have you provided a clear, succinct conclusion on the main findings on the REA question?	
19. Have you clearly described all limitations and discussed how they may impact on the findings of the REA?	
20. Have you clearly described what the implications for practice are?	



## References

- Antman, E. M. (1992). A comparison of results of meta-analyses of randomized controlled trials and recommendations of clinical experts. *Journal of the American Medical Association*, 286(2), pp. 240-248.
- Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76(2), 451-469.
- Bushman, B., & Wells, G. (2001). Narrative impressions of literature: The availability bias and corrective properties of meta-analytic approaches. *Personality and Social Psychology Bulletin*, 27(9), 1123-1130.
- Chalmers, I., Enkin, M., & Keirse, M. (1993). Preparing and updating systematic reviews of randomized controlled trials of health care. *Millbank Quarterly* (71), 411-437.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cooper, H. M., & Lindsay, J. L. L. (1998). *Research synthesis and meta-analysis*. p. 271-285. Sage Publications.
- Denyer, D (2014) The stages of a systematic review. Retrieved from <http://www.restore.ac.uk/logicofenquiry/logicofenquiry/gst/SR/stages/Pages/default.html>
- Fink, A. (1998). *Conducting Research Literature Reviews: From Paper to the Internet*. London: Sage Publications.
- Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews*. Sage.
- Government Social Research Service (2009) *GSR rapid evidence assessment toolkit* (Online), [www.civilservice.gov.uk/my-civil-service/networks/professional/gsr/resources/gsr-rapid-evidence-assessment-toolkit.aspx](http://www.civilservice.gov.uk/my-civil-service/networks/professional/gsr/resources/gsr-rapid-evidence-assessment-toolkit.aspx)
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23.
- Higgins, J., & Green, S. (Eds.). (2006). *Cochrane Handbook for Systematic Reviews of Interventions* (Vol. Version 5.1.0. (updated March 2011)). Chichester, UK: John Wiley & Sons, Ltd.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology*, 1(1)
- Newman M, Elbourne D (2005) Improving the usability of educational research: guidelines for the reporting of primary empirical research Studies in Education (the REPOSE guidelines). *Evaluation and Research in Education* 18 (4), 201-212.
- Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *British Medical Journal*, 322(January), 98-101.
- Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Ziliak, S. (2016). Statisticians Found One Thing They Can Agree On: It's Time To Stop Misusing P-Values. Retrieved January 04, 2017, from <http://fivethirtyeight.com/features/statisticians-found-one-thing-they-can-agree-on-its-time-to-stop-misusing-p-values/>

## Appendix: Summarizing scientific literature

Evidence summaries come in many forms. One of the best-known types is the conventional literature review, which provides an overview of the relevant literature published on a topic. However, a conventional literature review's trustworthiness is often low: clear criteria for inclusion are often lacking (including whether cited work is peer reviewed), studies are selected based on the researcher's individual preferences, and the research methodology is generally not subjected to a critical appraisal (Antman, 1992; Bushman & Wells, 2001; Chalmers, Enkin, & Keirse, 1993; Fink, 1998). As a result, most conventional literature reviews are prone to severe bias and are therefore largely considered untrustworthy as an answer to questions relevant to practice. For this reason, many evidence-based disciplines use so-called 'systematic reviews' instead. This type of review is a specific methodology that aims to comprehensively identify all relevant studies on a specific topic, and to select appropriate studies based on explicit criteria. In addition, the methodological quality of the studies included is assessed by on the basis of explicit criteria, such as the presence of a pre-test or a control group (Higgins & Green, 2006; Petticrew & Roberts, 2006). In contrast to a conventional literature review, a systematic review (SR) is transparent, verifiable, and reproducible, and, as a result, the likelihood of bias is considerably smaller. Many SRs also include a meta-analysis, in which statistical analysis techniques are used to combine the results of individual studies to arrive at a more accurate estimate of effects<sup>2</sup>.

Although the SR methodology was originally developed in the field of medicine, its added value is evident in disciplines such as nursing, education, policing, criminology, public policy, and management (Petticrew, 2001). In disciplines in which evidence-based practice is well established, SRs are provided by global communities such as the Cochrane and Campbell collaborations, and by organizations such as the EPPI Centre. In management, however, the SR methodology is not yet widely adopted, and systematic reviews are consequently scarce.

Rapid Evidence Assessments (REAs) and Critically Appraised Topics (CATs) are two other types of evidence summaries that can inform practice. Both apply the same systematic approach to selecting the studies – the methodological quality and practical relevance of the studies are assessed based on explicit criteria; thus, summaries are transparent, verifiable and reproducible. The main way in which these three types of summaries vary is in relation to the time and resources used to produce them and the scope and depth of the results produced. CATs are the quickest to produce and may take one skilled person a few days to produce. REAs might take two skilled persons several days to weeks. A SR usually takes a team of academics several months to produce, as it aims to identify all published and unpublished relevant studies (see table on the next page).

In general, an organization will not have time or financial means to hire a team to conduct a SR on a managerial topic of interest. A CAT, on the other hand, may be a good way to get a quick impression of the available scientific evidence regarding the effect of a specific intervention, but it may be lacking rigor. As a result of these practical limitations, an REA is the most widely used method of reviewing the scientific literature within Evidence-Based Management.

---

<sup>2</sup> The difference between a systematic review and a meta-analysis is not always clear. Many studies defined as systematic reviews include a meta-analysis aggregating statistical effects; conversely, studies defined as meta-analyses often include a systematic review of the literature.

Quality criteria	SR	REA	CAT	LR
The search for studies is systematic and reproducible	✓	✓	✓	X
All relevant research databases are searched	✓	✓	X	X
Attempts are made to locate unpublished research	✓	✓/X	X	X
Studies are selected based on explicit inclusion and exclusion criteria	✓	✓	✓	X
The study selection is conducted by two reviewers, independently of each other	✓	✓	X	X
The selection process is clearly documented, for example, in the form of a flow chart that shows how many studies are excluded and why	✓	✓	X	X
The process to extract data are clearly defined and is presented in a table	✓	✓	X	X
The methodological appropriateness of each study included is appraised using predetermined quality criteria	✓	✓	✓	X
The methodological quality of each study included is appraised using predetermined quality criteria	✓	✓	X	X
The appraisal of the methodological quality is conducted by two reviewers, independently of each other	✓	✓/X	X	X
The practical and methodological heterogeneity of the studies is assessed.	✓	✓	X	X
The statistical heterogeneity of the studies is assessed.	✓	X	X	X
Statistical analysis techniques (meta-analyses) are used to combine the results of individual studies to arrive at a more accurate estimate of effects	✓/X	X	X	X



**CEBMa**  
*center for*  
**Evidence-Based Management**

Center for Evidence Based Management  
[www.cebma.org](http://www.cebma.org)  
Amsterdam



ISBN 978-90-826957-0-0



9 789082 695700

