CEBMa
center for
Evidence-Based Management

CEBMa Guideline for

Critically Appraised Topics

in Management and Organizations

Version 2.0

Contents

For further information on CEBMa, visit our website: www.cebma.org

Critically Appraised Topics

A Critically Appraised Topic (CAT) provides a quick and succinct 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 primary studies. However, in order to be quick, a CAT makes concessions in relation to the breadth, depth and comprehensiveness of the search. Aspects of the search may be limited to produce a quicker result:

- **Searching:** a limited number of databases may be consulted, and unpublished research can be excluded. Sometimes a CAT may be limited to only meta-analyses and/or controlled studies.

- **Data Extraction:** only a limited amount of key data may be extracted, such as year, population, sector, sample size, main findings, and effect size.

- **Critical Appraisal:** quality appraisal is often limited to methodological appropriateness.

Due to these limitations, a CAT is more prone to selection bias than a systematic review or rapid evidence assessment.

Steps in the CAT process

A CAT 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
9. Conclusion
10. Limitations
11. Implications and recommendations

Step 1. Background: What is the context of the CAT question

The background should clearly state what the rationale for the CAT 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. Most CATs occur in the context of a specific organization. You should address this context (e.g. sector, history, characteristics), help specify the rationale for the CAT, and explain why the question is important for the organization, its members or its clients. In formulating the CAT 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

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. Particularly when the change has predominantly negative outcomes (e.g., lay-offs), I think it is of utmost importance that the change process is fair and just. I am curious about the impact procedural justice has on the way people perceive the outcomes of organizational change.

EXAMPLE 2

Interviewing and who got what job how are the most talked about subject on campus. Most students are getting ready to either get internships or full time jobs this summer. It is widely believed that smiling during a job interview may increase your chances of getting hired. This CAT was conducted to understand whether this claim is supported by scientific evidence.

EXAMPLE 3

Hoping to imitate the innovative and flexible work environments found at start-ups and companies like Google, my organization's Executive Director is considering implementing a bullpen style, open-plan layout. Currently, our office is divided into individual workspaces with half walls. The Executive Director and Controller both have private offices. With 15 employees working in a relatively small space, I worry that the distractions created by a new, open layout may undermine our ability to focus and to be productive at work. To draw a more informed conclusion on the effect that such a layout might have at our office, I have gathered and assessed the quality of available scientific evidence, outlined key findings, and summarized their practical implications.

Step 2. Formulating the CAT question: What does the CAT answer?

You can use a CAT 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 CAT question are about

- the effect an intervention, factor, or independent variable.

- the drivers (antecedents) of a certain outcome.

EXAMPLE 1

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

EXAMPLE 2

What is known in scientific literature about the impact of smiling during job interviews: Do people who smile more have better chances of getting hired?

EXAMPLE 3

What is known in the scientific literature about the impact of flexible work schedules on task performance?

EXAMPLE 4

What is known in the scientific literature about the effect of open-office layouts on workers' task performance?

Non-impact questions

You can, however, use CATs 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:

Population	Who?	Type of employee, subgroup, people who may be affected by the outcome
Intervention	What or How?	Management technique/method, factor, independent variable
Comparison	Compared to what?	Alternative intervention, factor, variable
Outcome	What are you trying to accomplish / improve / change?	Purpose, objective, dependent variable
Context	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 a CAT 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 CAT when reviewing its abstract and/ or full text. The inclusion criteria should be guided by your CAT 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:** only meta-analyses or controlled studies

5. **Measurement:** a) studies in which the effect of goal setting on organizational outcomes was measured or b) studies in which the effect of moderators and/or mediators on the outcome of goal setting was 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.

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 CAT 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 two databases were used to identify studies: ABI/INFORM Global and PsycINFO. The following generic search filters were applied to all databases during the search:

1. Scholarly journals, peer-reviewed

2. Published in the period 1990 to 2020 for meta-analyses and the period 2010 to 2020 for primary studies

3. Articles in English

A search was conducted using combinations of different search terms, such as 'job interview', 'employment interview', 'selection interview' and 'smiling'. An overview of all search terms and queries is provided in Appendix I.

Documentation of the search

It is important that the search conducted is transparent, verifiable and reproducible. For this reason, the search process should be clearly documented, preferably in the form of a table that shows which search terms were used, how search terms were combined, and how many studies were found at every step. An example is provided below.

ABI/Inform Global, PsycINFO peer reviewed, scholarly journals, July 2020		
Search terms	**ABI**	**PSY**
S1: ti("job interview*") OR ab("job interview*")	76	231
S2: ti("employment interview*") OR ab("employment interview*")	122	368
S3: ti("selection interview*") OR ab("selection interview*")	70	130
S4: S1 OR S2 OR S3	259	583
S5: ti(smil*) OR ab(smil*)	736	2,673
S6: S4 AND S5	7	5

Note

For the sake of readability, your overview of search terms, queries, and results should be kept in a separate appendix.

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

In general, a search will yield a large number of studies – sometimes more than hundred. Some studies will not be directly relevant to the CAT question. Hence, the next step is to screen them to check that they meet the inclusion criteria. Screening is usually a two-stage process, the first involves reviewing the titles & abstracts and the second, reviewing the full studies.

Review titles & abstracts

This involves reading the titles abstracts that have been found through the searching. If the study seems relevant then the study should be included. Unfortunately, not all abstracts will contain sufficient information to determine whether the study is relevant. When in doubt, the study should be included.

Review full studies

Next, you should 'scan' the full article and compare it against the inclusion criteria.

EXAMPLE

Selection took place in two phases. First, the titles and abstracts of the 126 studies identified were screened for their relevance to this CAT. In case of doubt, the study was included. Duplicate publications were removed. This first phase yielded 86 studies. Second, studies were selected based on the full text of the article and compared against the inclusion criteria. This second phase yielded a total number of 8 studies.

Step 6. Data extraction: What information should be extracted?

Data extraction involves the collation of the results and other information of the studies included. From each study, information relevant to the CAT question, such as year of publication, research design, sample size, population (e.g., industry, type of employees), outcome measures, main findings, effect sizes, weaknesses, and the final level of trustworthiness (see 7) should be reported, preferably in the form of a clearly structured table. An example is provided on the next page.

Note

For the sake of readability, your data-extraction table hould be kept in a separate appendix.

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	Worksite; 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	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 trail, 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	Small	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	Moderate	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	Artificial students	**A**

Step 7. Critical appraisal: How should the quality of the studies be judged?

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[1].

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

---

[1] 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 CAT 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).

EXAMPLE 1

The overall quality of the studies included was high. Of the 4 meta-analyses, 3 included randomized and/or non-randomized controlled studies and were therefore qualified as level A or AA. The remaining meta-analysis was graded as level C, because it was insufficiently clear what type of studies were included. The actual level of evidence of this meta-analysis (and as result the overall quality of the studies included in this CAT) may therefore be higher. All 3 primary studies used a cross-sectional design and were therefore graded level D.

EXAMPLE 2

After critical appraisal of the 24 studies, only four studies were included. Most studies were excluded because they had serious methodological shortcomings. One of the studies included concerned 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 between studies it was not possible to calculate a pooled estimate of effect. The three single primary studies used a cross-sectional design. As a result, the trustworthiness of the scientific evidence supporting the following main findings is very limited.

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, Δ, g	≤ .20	.50	≥ .80
Correlation: r, ρ	≤ .10	.30	≥ .50
Correlation: r^2	≤ .01	.09	≥ .25
ANOVA: η^2, ω^2	≤ .01	.06	≥ .14
Chi-square: ω^2	≤ .10	.30	≥ .50
Simple regression: β	≤ .10	.30	≥ .50
Multiple regression: β	≤ .20	.50	≥ .80
Multiple regression: R^2	≤ .02	.13	≥ .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 that was measured.

Step 8. Results: What did you find?

Step 8.1. Definition: What is meant by X?

Most CAT 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" (Doe, 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 (Doe, 2014). Formally, incentives differ from rewards. Incentives refer to all stimuli that are provided in advance, whereas rewards are offered after a given performance (Doe, 2014). In the scientific literature and management practice, however, these terms are used interchangeably.

EXAMPLE 2

A smile is defined as a pleased, kind, or amused expression, typically with the corners of the mouth turned up and the front teeth exposed. A neutral expression is a blank facial expression characterized by neutral positioning of the facial features, implying a lack of strong emotion.

EXAMPLE 3

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. Doe et al (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" (Doe, 2005: p. 341).

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 presumed causal mechanism behind open-plan offices works as follows:

- *Breaking down physical office barriers increases one's exposure to teammates*
- *Increased exposure facilitates communication among teammates*
- *Facilitated communication increases collaboration*
- *Increased collaboration increases overall productivity and performance*

However, one can also presume that increased exposure to teammates might also cause increases in visual and auditory distractions, counteracting positive benefits to some degree.

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 (Doe et al., 2013; Doe, 1960). Put differently, when a manager helps his or her employees in times of need or recognizes their extra effort, the employees will feel inclined to act to benefit that manager (e.g., by meeting goals and objectives) and thus the organization as a whole (Doe, 2013; Doe et al, 1986). Not surprisingly, physicians experiencing inadequate supervisory support tend to describe their commitment to the hospital and its patients in negative terms, which in turn negatively affects their performance (Doe et al, 2013).

Step 8.3. Main findings

In this section, you should provide an overview of the main findings relevant to the CAT question. For each finding, you should present the main evidence from the CAT, including its level of trustworthiness and (if available) effect size. Report each finding in the form of a brief statement and the level of evidence, followed by a short explanation and a reference to the study(s).

EXAMPLE 1

Smiling is weakly correlated with job interview success (level D)

Putting on a smile may be advantageous by comparison with remaining neutral, which may be seen as reflecting a lack of interest or involvement. But there are other factors at play too during the interview and just smiling is weakly correlated to success (Doe et al, 2009).

EXAMPLE 2

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 (Doe, et al., 2014; Doe & Doe, 2014; Doe, 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, as described in the following findings. Moreover, when financial incentives are high they can hamper performance by interfering with learning and promoting inappropriate risk taking (Doe et al., 2009).

EXAMPLE 3

When employees must first acquire requisite knowledge or skills to perform the task, specific and challenging goals can have a large negative effect on performance (level A)

Several randomized controlled studies have demonstrated that when a task requires the acquisition of knowledge before it can be performed effectively, a general goal (e.g. 'do your best') leads to higher performance than a specific high goal (Doe & Doe, 1989; Doe, 1995). In fact, when knowledge acquisition is necessary for effectively performing a task, setting a specific but extremely high performance goal can lead people to ruminate on the potential negative consequences of failure rather than focus on task-relevant ways to attain the goal (Doe et al, 2002).

EXAMPLE 4

Arousing fear leads employees to engage in short-term thinking (level B)

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, narrowing their attention to immediate consequences (Doe, 1999).

EXAMPLE 5

The outcome of the managers' own performance appraisal has a large effect on how they evaluate their employees (level A)

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

EXAMPLE 6

Managers' power level has a large to moderate effect on how they rate both others and themselves (level A)

A meta-analysis of 46 studies indicates that as a manager's power level grows, his/her evaluation of others becomes increasingly negative, whereas evaluations of him/herself become ever more positive (Doe & Doe, 1998). This finding suggests that performance evaluations by supervisors should be considered in light of their hierarchical position and power level.

EXAMPLE 7

Managing the customer experience may have a moderate positive effect on a company's financial performance (level D)

Findings from cross-sectional studies suggest that proactive customer experience management (CEM), such as customer touch points, customer focus, goals for customer experiences, customer insight, employee recruitment and training, etc., may have a positive effect on a company's financial performance (Doe, 2015).

Step 9.   Conclusion

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

EXAMPLE 1

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 during organizational change.

EXAMPLE 2

Corporate social responsibility has been shown to have a positive relation with corporate financial performance. Although a causal could not be demonstrated, there is a bi-directional correlation between the two variables, indicating that corporate social responsibility is correlated with financial performance and vice versa. Furthermore, while firm size and year of observation have been shown not to have an impact, other confounding variables may still affect the relationship.

EXAMPLE 3

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 4

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

Step 10. **Limitations**

In a CAT 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 CAT have limitations. In your CAT 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 CATs.

EXAMPLE

Concessions were made in relation to 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 CAT did not conduct 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 CAT focused only on meta-analyses. As a consequence, relevant findings may have been missed.

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

Step 11. **Implications and recommendations**

Once you have used the evidence found to answer the CAT's main question, you should use the final part of the assessment to relate the findings to the background of the CAT 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 aspects of their decision-making process, specifically consistency, accuracy, lack of bias, and openness to employee input. When procedural justice is not taken into account, employees may feel treated unfairly and resistance to change may increase. To actively design a fair change process, the six classic criteria for procedural justice specified by Doe (1980) 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) opportunities should be provided to correct or change plans or processes; (e) those responsible for the organizational change (the change managers or leaders) 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 to support change. 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 and quality outcomes should be encouraged by indirect incentives (e.g., opportunities to do valued activities) in order to avoid eroding that motivation. Direct financial incentives are effective when extrinsic motivation and quantitative performance need to be stimulated. Upper management should therefore frequently 'calculate' the proposed net effect (positive price effect versus negative crowding-out effect) when defining a pay plan. Lastly, if the plan is designed to increase team performance, all incentives should not be distributed equally, as this may harm individual motivation.

EXAMPLE 3

The continuation of the client's mentoring program should be seriously weighed at this time, given the evidence that (a) the effect of youth mentoring on academic outcomes is small and (b) fidelity to matching relationships and sustaining them over the long term may provide greater benefit. Given that the organization's program is school-based and effective long-term mentors are challenging to recruit and retain, the organization may want to consider suspending its mentor program. Given that the organization's staff is overextended and that budget allocations across programs are at-risk, I would recommend that mentoring activities be suspended at the end of the school year.

EXAMPLE 4

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. Neither can it be determined whether self-managing teams will help our organization to change more successfully. Instead, we suggest that our organization should be extra careful about implementing self-managing teams when the divisions involved have high levels of hierarchy, centralized decision making or bureaucratic formalization. Divisions characterized by high-tech novelty and radical innovation, however, do not need to be as cautious in implementing self-managing teams. The latter 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 should therefore be considered carefully.

EXAMPLE 5

This CAT demonstrates that Emotional Intelligence (EI) is not a radical new construct in 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 CAT. For this reason, I advise against investing in training courses that claim to develop our executives' EI.

EXAMPLE 6

Over the last 20 years, there has been a huge increase in consumer awareness, tighter environmental regulations, and a shift towards more environmentally friendly business practices. This review demonstrates that corporate social responsibility (CSR) strategies are both required to mitigate environmental and regulatory risk and correlate with higher financial performance. This relationship is recognized regardless of industry, firm size, or the year of study. Therefore, I would recommend that managers explore and assess opportunities to unlock value from CSR strategies throughout business operations.

Checklist

1. Have you clearly described the background and context of the CAT question?	
2. Does the CAT address a clearly focused question? Is it clear what the CAT will answer?	
3. Have you used the PICOC framework to focus the CAT question?	
4. Have you clearly defined the inclusion criteria (e.g. population, outcomes of interest, study design)?	
5. Have you conducted a comprehensive literature search using relevant research databases (i.e. ABI/INFORM, Business Source Premier, PsycINFO, Web of Science, etc.)?	
6. Is the search systematic and reproducible (e.g. were searched information sources listed, were search terms provided, were search results reported)?	
7. Have you selected the studies using explicit inclusion and exclusion criteria?	
8. 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?	
9. Have you assessed the methodological appropriateness of each study using predetermined quality criteria?	
10. Have you provided definitions of the key elements/constructs in the CAT question?	
11. Have you clearly described the assumed causal mechanism?	
12. Have you provided an overview with the main findings, including their level of trustworthiness and effect size?	
13. Have you provided a clear, succinct conclusion on the main findings on the CAT question?	
14. Have you clearly described all limitations and discussed how they may impact on the findings of the CAT?	
15. 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

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[2].

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.
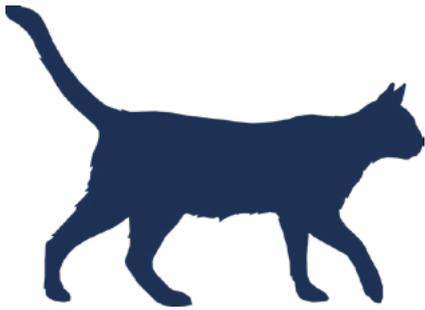
---

[2] 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.