CEBMa
center for
Evidence-Based Management

Open Learning Initiative

Evidence-based Management

How to use evidence to make better organizational decisions

TEXT CERTIFIED COURSE - 6th EDITION

PLEASE NOTE:

This document contains the text of the 15 modules of CEBMa's certified online course "*Evidence Based Management - How to use evidence to make better organizational decisions*". It can be used only in combination with the online course and is not meant as a stand alone document. Its content is protected by copyright and is the property of the Center for Evidence Based Management.

This means you may NOT copy, distribute, share, or publish its content.

Center for Evidence Based Management
Leiden, The Netherlands, 2025

Pubished by the Center for Evidence-Based Management, Leiden, The Netherlands

ISBN: 978-90-826957-62

CEBMa center for Evidence-Based Management

Open Learning Initiative

Table of contents

CEBMa center for Evidence-Based Management

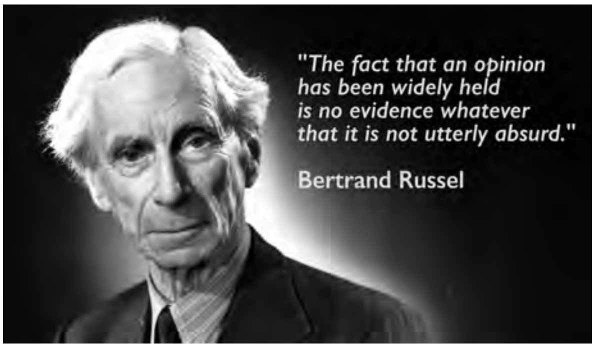
Open Learning Initiative

MODULE 1 | The Basic Principles of Evidence-based Management

Learning objectives:

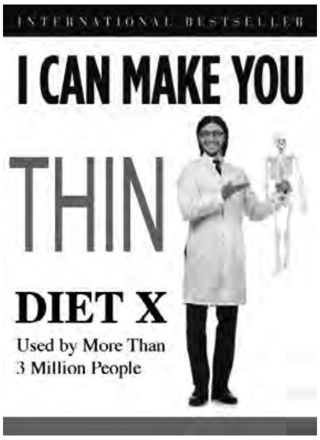
- Summarize the basic principles of evidence-based management.
- Explain the need for evidence-based management.
- Identify the four sources of evidence, and provide examples of each.
- Evaluate (coarsely and in general terms) the quality of evidence.
- Determine whether the best available evidence was used in a decision-making process.
- Correct common misconceptions about evidence-based management.

Introduction



Consider this hypothetical situation

You pay a visit to a dietitian after gaining a bit of weight during the holiday season. The dietitian advises you to try diet X. It's very expensive and demands a radical change in lifestyle, but the prospect of having a slim and healthy body motivates you to stick to the diet. After a few weeks, however, you have gained five pounds and suffer from serious side effects that require medical treatment. By searching the Internet, you learn that most scientific studies find diet X to be ineffective and fraught with side effects. When you confront the dietitian with these findings, he answers, "Why should I pay attention to scientific studies? I have twenty years of experience. Besides, the diet is developed by a famous American nutritionist, whose book sold more than a million copies." [1]



Does that scenario sound like malpractice to you? It probably does. Unfortunately, in management disregarding sound evidence and relying on personal experience or the popular ideas of management gurus is daily practice. Decisions made by managers, however, affect the working lives and well-being of people around the world. As Henry Mintzberg said: "No job is more vital to our society than that of a manager. It is the manager who determines whether our social institutions serve us well or whether they squander our talents and resources." [2]

In this module, we explain what evidence-based practice is and how it can help you and your organization to make better decisions. Whether we work in a banking firm, hospital, large consulting firm, or small startup, as practitioners affecting the lives of so many, we also have a moral obligation to use the best available evidence when making a decision. We can do that by learning how to distinguish science from folklore, data from assertions, and evidence from beliefs, anecdotes, or personal opinion.

What is Evidence-based Management?

The basic idea of evidence-based management is that sound quality decisions require both critical thinking and use of the best available evidence. Of course, all practitioners use evidence in their decisions. But few pay attention to the *quality* of the evidence and tend to base their decisions on only one source. The result is decisions that rely on unfounded beliefs, fads and fashions, and the unsupported though popular ideas of management gurus. The bottom line is bad decisions, poor outcomes, and little understanding of why things go wrong.

Evidence-based management seeks to improve the way decisions are made. It is an approach to decision-making and day-to-day work practice that helps practitioners to critically evaluate the extent to which they can trust the evidence they have at hand. It also helps practitioners identify and evaluate additional evidence relevant to their decisions. In this course, we use the following definition of evidence-based management. This definition is adapted in part from the Sicily statement of evidence-based practice. [3] It not only provides a clear statement of what evidence-based management means but also describes the main skills required to practice in an evidence-based way:

Evidence-based management is about making decisions through the conscientious, explicit, judicious use of the best available evidence from multiple sources by . . .

ASKING:	Translating a practical issue or problem into an answerable question
ACQUIRING:	Systematically searching for and retrieving the evidence
APPRAISING:	Critically judging the trustworthiness and relevance of the evidence
AGGREGATING:	Weighing and pulling together the evidence
APPLYING:	Incorporating the evidence into the decision-making process
ASSESSING:	Evaluating the outcome of the decision taken

. . . to increase the likelihood of a favorable outcome.

Learn by doing 1.1

Take another look at the definition of evidence-based management. As you can see, it is quite detailed. If you had to explain to someone what evidence-based management is, which two element(s) would you emphasize?

a) Evidence-based management is about assessing the outcome of the decisions you take.

b) Evidence-based management is about taking a systematic (six-step) approach to decision-making.

c) Evidence-based management is about basing your decisions on critically appraised evidence.

d) Evidence-based management is about basing your decisions on multiple sources of evidence.

What counts as evidence?

When we say *evidence*, we basically mean **information**: facts or data supporting (or contradicting) a claim, assumption, or hypothesis. It may be based on numbers, or it may be qualitative or descriptive. Evidence may come from scientific research suggesting some relatively generalizable facts about the world, people, or organizational practices. Evidence may also come from local organizational or business indicators, such as company metrics or observations of practice conditions. Even professional experience can be an important source of evidence, such as when an entrepreneur learns from having launched a variety of businesses that one particular approach seems more likely to pay off.

Think of it in legal terms. In a court of law, evidence from many different sources is presented, including eyewitness testimony, forensic evidence, security camera images, and witness statements. All this evidence may help a judge or a jury to decide whether a person is innocent or guilty. The same is true for management decisions. Regardless of its source, all evidence may be included if it is judged to be trustworthy and relevant.

Learn by doing 1.2

1. Imagine you are an executive of a large corporation. You are thinking about introducing flexible working arrangements. Would you regard the experience of employees with flexible working arrangements as evidence?
2. Imagine you are a senior manager of a large organization. Would you regard the following statement from a business consultant as evidence of how long it takes to change an organization's culture?

   "It is common knowledge that it takes years to change an organization's culture."
3. Imagine you are a manager at an IT firm. Would you regard the following statement from a colleague as evidence about information sharing between software engineers?

   "In the 15 years I have worked as a manager, I have noticed that software engineers are more likely to share information when they trust each other."

Did I get this 1.1?

Would you regard the following statement as evidence about how to change employees' behavior?

"According to our CEO, who has a degree from Stanford, you must first change the attitudes of people in order to change their behavior."

Why do we need Evidence-based Management?

Learn by doing 1.3

Trust me,
I have 20 years of management experience

Practitioners use many sources of evidence when making decisions, including intuition, personal experience, knowledge acquired through formal education, insights provided by experts, advice from a colleague, management literature, scientific research, and others.

On what do you base your decisions?

Most management decisions are not based on the best available evidence. Instead, practitioners often prefer to base decisions solely on their judgment derived from personal experience. However, personal experience alone is not a reliable source of evidence because it is highly susceptible to systematic errors. We have cognitive and information-processing limits that make us prone to biases that have negative effects on the quality of the decisions we make. [4] [5] [6] [7]

Even practitioners and industry experts with many years of experience perform poorly when making forecasts or calculating risks when relying solely on their personal judgment, whether it concerns the credit rating of bonds, [8] the growth of the economy, [9] political developments, [10] or medical diagnoses. [11]

Another heavily used source of evidence seems to be what other organizations are doing. Through benchmarking and so-called best practices, practitioners sometimes copy the methods and procedures of other organizations without critically evaluating whether those practices are actually effective and, if they are, whether they are also likely to work in a different context. Benchmarking can demonstrate alternative practices, but it is not necessarily a good indicator in itself of what would work in a different setting.

At the same time, there are many barriers to evidence-based practice. Few practitioners have been trained in the skills required to critically evaluate the trustworthiness and relevance of the information they use. In addition, important organizational information may be difficult to access, and what is available can be of poor quality. Finally, practitioners are often unaware of the current scientific evidence available on key issues in the field. Fo example, a survey of 950 American HR practitioners showed large discrepancies between what practitioners think is effective and what the current scientific research shows. [12] This study has been repeated in other countries with similar findings. More educated managers do, however, show somewhat greater knowledge of scientific findings.

These results suggest that most practitioners pay little or no attention to scientific or organizational evidence. Instead, the typical practitioner seems to place too much trust in low-quality evidence such as personal judgment and experience, best practices, and the beliefs of corporate leaders. As a result, billions of dollars are spent on management practices that are ineffective or even harmful to organizations, their members, and their clients.

Example

For years, the US technology company Google believed that technical expertise was the most important capability for their managers. The company's leaders thought that the best managers left their people alone as much as possible, focusing instead on helping them with technical problems when people got stuck. When the company examined what employees valued most in a manager, however, technical expertise ranked last among eight qualities. More crucial were attributes such as asking good questions, taking time to meet employees, and caring about employees' careers and lives. Managers with these qualities led top-performing teams and had the happiest employees and lowest turnover. These attributes of effective managers, however, also are well established in scientific studies, so Google s improvement efforts could have started years earlier.

Learn by doing 1.4

Evidence-based management is about making decisions based on the best available evidence, that is, *critically appraised* evidence from *multiple sources*.

Why do we need evidence-based management? Check the three most important arguments.

a) Management practitioners often base their decisions solely on their judgment derived from personal experience. However, personal judgment alone is not a reliable source of evidence.

b) Good-quality evidence from multiple sources gives you the answer to most organizational problems.

c) Through benchmarking and so-called best practices, managers sometimes copy the methods and conventions of other organizations without critically evaluating whether these practices are actually effective.

d) There are large discrepancies between what management practitioners think is effective and what the current scientific research shows.

To give evidence-based management the best chance at success, we need to increase the capacity of managers and organizations to prioritize quality evidence over unfounded personal opinion and to incorporate what the body of evidence indicates into their better-informed professional judgment.

What sources of evidence should be considered? (1)

Before making an important decision, an evidence-based practitioner starts by asking, “What is the available evidence?” Instead of basing a decision on personal judgment alone, an evidence-based practitioner finds out ***what is known*** by looking for evidence from multiple sources.

According to the principles of evidence-based practice, evidence from four sources should be taken into account:

- **Evidence from the scientific literature:** Findings from published empirical studies
- **Evidence from the organization:** Data, facts, and figu es gathered from the organization
- **Evidence from practitioners:** The professional experience and judgment of practitioners
- **Evidence from stakeholders:** The values and concerns of people who may be affected by the decision

Evidence from the Scientific Literature

The first source of evidence is published research findings. When we use the term *scientific evidence*, we mean the evidence from scientific research published in academic journals. Over the past few decades, the volume of management research has escalated dramatically. Topics range from evaluating the success of mergers and the effects of financial incentives on performance to improving employee commitment and recruitment. In addition, much of the research from outside the management discipline is still highly relevant because many of the problems managers face are not so different from those people would experience in any situation – how to make better decisions, communicate more effectively, or manage conflict. When it comes to tackling these issues in management practice, it makes sense to consider relevant scientific findings om any discipline. Although many practitioners learn about research findings in their education or professional courses, new research findings a e produced regularly, and these findings can often change our understanding. To include up-to-date scientific evidence in your decisions, it helps to know how to search for studies and to be able to judge how trustworthy and relevant they are.

Example

The board of directors of a large Canadian law firm has plans for a merger with a smaller firm nearby The merger’s objective is to integrate the back offices of the two firms (IT, finance, facilities, etc.) in order to create economies of scale. The front offices and legal practices of the two firms will remain separate. The board has been told by the partners that the organizational cultures of the two firms differ widely, so the board wants to know whether the differences are likely to create problems for the merger. Partners of both firms we e asked independently about their professional experience with mergers. Those who had been involved in one or more mergers stated that cultural differences matter and can cause serious culture clashes between professionals.

How did scientific evidence help?

A search conducted in online scientific databases yielded a meta-analysis based on 46 studies with a combined sample size of 10,710 mergers and acquisitions. The meta-analysis confirms the partner s judgment that there is a negative association between cultural differences and the effectiveness of the post-merger integration. However, the study also indicates that this is only the case when the intended level of integration is high. In mergers that require a low level of integration, cultural differences are found to be positively associated with integration benefits. In the case of the two law firms, the planned integration concerns only back-office functions, making the likelihood of a positive outcome higher.

In Module 5, ACQUIRE Evidence from the Scientific Literature, you will learn the skills necessary to successfully search for relevant empirical studies using online research databases such as ABI/INFORM Global, Business Source Premier, and PsycINFO.

Evidence from the Organization

A second source of evidence is the organization itself. Evidence from the organization, whether a business, a hospital, a governmental agency, or a volunteer organization, comes in many forms. It can be financial data, such as cash flow or cost, or business outcomes, such as return on investment or market share. It can come from customers or clients in the form of customer satisfaction, repeat business, or customer recommendations. It can also come from employees through information about retention rates or levels of job satisfaction. Organizational evidence can consist of hard numbers such as turnover rates, medical errors, or productivity levels, but it can also include soft elements, such as perceptions of the organization's culture or attitudes toward the senior leaders. Knowing the organizational evidence is essential in deciding whether a problem exists that requires managers' attention. It is also essential in identifying the likely causes of a problem, plausible solutions, and what is needed to implement these solutions.

Example

The board of a large insurance company plans to change its regional structure to a product-based structure. According to the board, the restructuring will secure the company's market presence and drive greater customer focus. The company's sales managers strongly disagree with this change, arguing that ditching the region-based structure will make it harder to build good relationships with customers and will therefore harm customer service.

How did evidence from the organization help?

Analysis of organizational data revealed that the company's customer satisfaction is well above the industry average. Further data analysis showed a strong negative correlation between the account managers' monthly travel expenses and the satisfaction of their customers, suggesting that sales managers who live close to their customers score higher on customer satisfaction.

In Module 9, APPRAISE Evidence from the Organization, you will develop a better understanding of evidence from the organization and learn to acquire it in a valid and reliable way.

Learn by doing 1.5

1. How would you define evidence from the scientific literature (in one or two sentences)?
2. How would you define evidence from the organization (in one or two sentences)?

Evidence from Practitioners

A third source of evidence is the professional experience and judgment of managers, consultants, employees, staff, and other practitioners. Different from intuition, opinion, or belief, professional experience is accumulated over time through the feedback received on the outcomes of similar actions taken in similar situations. This type of evidence is sometimes referred to as *tacit knowledge*. Professional experience differs from intuition and personal opinion because it reflects the specialized knowledge or expertise acquired by repeated experience and practice of technical activities, such as playing the violin or making a cost estimate.

Many practitioners take seriously the need to learn thoughtfully and critically from their experiences. Their knowledge can be vital for determining whether a management issue really does require attention or whether the organizational data available are trustworthy. Thoughtful practitioners also use their experience to judge whether research findings apply in a particular situation or a proposed solution is likely to work in a particular context. If relevant and trustworthy, experiential evidence plays a key role in the decision-making process.

Example

A Dutch university hospital has decided to implement personal development plans for all its nurses. These plans would include a statement of the nurse's aspirations and career priorities. The HR director points out that according to Maslow's hierarchy of needs, a well-known motivation theory, basic levels of needs (such as health and safety) must be met before an individual can focus on his or her higher-level needs (such as career and professional development). The nurses at the emergency department are increasingly exposed to serious safety hazards, from offensive language to physical violence. The HR director therefore recommends excluding these nurses from composing a personal development plan until the safety hazards are under control and significantly reduced.

How did evidence from practitioners help?

Experienced managers and nurses were asked independently about their view on the director's recommendation. Most of them disagreed and indicated that their in their professional experience, the opposite often is true: nurses who work in difficult circumstances tend to be strongly interested in professional development and self-improvement. In addition, a search was conducted in online scientific databases. It yielded a range of studies indicating that the e is no empirical evidence available that supports Maslow's theory; consequently, the managers' and nurses' experience is supported.

In Module 3, ACQUIRE Evidence from Practitioners, we explain how to gather evidence from practitioners in a valid and reliable way, covering aspects such as what, who, and how to ask; the sample size needed; and how to develop appropriate questionnaires.

Evidence from Stakeholders

A fourth source of evidence is stakeholder values and concerns. Stakeholders are any individuals or groups who may be affected by an organization's decisions and their consequences. Internal stakeholders include employees, managers, and board members. Stakeholders outside the organization, such as suppliers, customers, shareholders, the government, and the public at large, may also be affected. Stakeholder values and concerns are a reflection of what stakeholders believe to be important, which in turn affects how they tend to react to a decision's possible consequences. Stakeholders may place more or less importance on, for example, short-term gain or long-term sustainability, employee well-being or employee output, organizational reputation or profitability, and participation in decision-making or top-down control.

Organizations that serve or respond to different stakeholders (e.g., ExxonMobil and Greenpeace) can reach very different decisions on the basis of the same evidence.

Gathering evidence from stakeholders is important not just for ethical reasons. Understanding stakeholder values and concerns also provides a frame of reference to make sense of and weigh evidence from other sources. It provides important information about the way in which decisions will be received and whether the outcomes of those decisions are likely to be successful.

Example

To assess employees' satisfaction with their supervisors, a British telecom organization conducted a survey among its 12,500 employees. The survey contained some demographic questions such as postal code, date of birth, and job title along with five questions on employee satisfaction with their immediate supervisor. The introductory letter by the CEO stated that all answers would remain anonymous. After the survey was sent out, only 582 employees responded, a response rate of less than 5 percent.

How did stakeholder evidence help?

A focus group discussion with employees from different parts of the organization was conducted to find out why so many members did not participate in the survey. The employees in the focus group stated that they were concerned that the demographic data would make it possible to identify the person behind the answers. Given the sensitive nature of the survey's topic, they therefore decided not to participate. Based on this outcome, the survey was modified by dropping the postal code and replacing the date of birth with an age range. The modified survey yielded a response rate of 67 percent.

Learn by doing 1.6

1. How would you define experiential evidence (in one or two sentences)
2. How would you define stakeholder evidence (in one or two sentences)

Did I get this 1.2

1. Imagine you are the CEO of a large chemical plant. Would you consider the concerns of the plant's surrounding community regarding chemical waste disposal, as reported by the local paper, as evidence?
2. Would you regard the outcome of the annual employee satisfaction survey as evidence?
3. Would you consider the concerns and fears employees have regarding a possible layoff to be evidence?

What sources of evidence should be considered? (2)

Example

Miracle on the Hudson

On January 15, 2009, Captain Chesley Sullenberger was flying out of LaGuardia Airport in New York City when his US Airways plane was struck by a large flock of geese. The bi d strike, which occurred only minutes after takeoff, damaged both engines and resulted in a rapid loss of power. With air traffic control, Sullenberger discussed his options: either return to LaGuardia or land at Teterboro Airport in New Jersey. Sullenberger quickly deemed the situation too dire for the plane to stay in the air long enough for either plan to be successful. Instead, he decided that ditching (performing an emergency water landing) the airplane in the middle of New York's Hudson River was the best option. He announced over the intercom, "Brace for impact," and took the plane down onto the water's surface. The maneuver was a success, and all 155 people on board survived. [13]

Would you consider Captain Sullenberger's decision to land the plane on the Hudson River as an evidence-based decision? Well, you might be surprised to find out that it was

Evidence from the Scientific Literature

After Sullenberger successfully landed his plane on the Hudson River, the flight crew immediately started an evacuation of the plane, getting people into rafts and out onto the wings. After all passengers and crew members had left the plane, Sullenberger inspected the plane very carefully to make sure no one was left on board. Sullenberger, however, was not quite satisfied. He had done research with NASA scientists on how to make decisions to maintain safety despite technological complexity and crisis conditions. From this NASA research and other research, Sullenberger knew that stress narrows the focus of attention and limits the extent to which new information can be absorbed. As he recalled: "I sensed my perceptional field was narrowed, you know, the tunnel vision that you get from sudden, life-threatening stress." For this reason, Captain Sullenberger went back and walked the entire length of the aircraft for a second time—to make sure he had not missed anyone.

Evidence from the Organization

While working for US Airlines, Sullenberger was an air accident investigator for the National Transportation Safety Board (NTSB). He analyzed flight data, cockpit voice recordings, and statistics from major incidents and wrote aviation accident reports for more than 20 years. From these data and statistics, Captain Sullenberger and his copilot, Jeffrey Skiles, knew it was critical to judge the height to begin raising the nose for landing; if the nose was raised too soon, the plane would lose lift on the wings and crash. For this reason, his copilot called out air speed and altitude to help Sullenberger judge the height in the final seconds before landing.

Evidence from Practitioners

Sullenberger served as a fighter jet pilot in the US Air Fo ce from 1975 to 1980 and was employed by US Airways from 1980 until 2010. In total, he achieved more than 20,000 flying hours. His years of professional experience helped Sullenberger to go quickly through the three pages of checklist procedures for the emergency landing and determine that the plane would not have enough power to make it back to LaGuardia Airport. In Sullenberger's own words: "*One way of looking at this might be that for 42 years, I've been making small, regular deposits in this bank of experience, education, and training. And on January 15, the balance was sufficient so that I could make a very large withdrawal.*"

Evidence from Stakeholders

The passengers were Sullenberger's most important stakeholders. Sullenberger realized that some of them, such as children and elderly people, may not be able to swim. For this reason, Sullenberger decided to ditch the plane on a part of the Hudson River that was near the ferry boat terminal. As a result, within five minutes after it hit the water, the plane was surrounded by ferry boats, and all 150 passengers were rescued.

Learn by doing 1.7

Read the following case and determine which sources of evidence were consulted.

A local branch of a British bank struggles with low employee satisfaction. Management has tried several initiatives to improve this situation, but this year's satisfaction survey has shown no improvement. After the outcomes of the survey became clear, senior management organized a meeting with all employees to ask them one basic question: What, in your opinion, is important for your job satisfaction? The outcome of this session indicated that the employees perceived trust in management as the most important factor.

1. Was scientific evidence consulted
2. Was organizational evidence considered?
3. Was professional experience and judgment taken into account?
4. Were stakeholder values and concerns considered?

Did I get this 1.3

Read the following case and determine which sources of evidence were consulted.

About 4 years ago, a large Belgian brewery (nine divisions, 6,800 employees) introduced Six Sigma, a management technique to decrease product defects and increase efficiency. However, despite Six Sigma being implemented in six divisions, results have declined over the past year, including:

- *The number of breakdowns and short stops in the production lines remains the same despite greater use of problem analysis and improvement effort.*
- *The average productivity per worker has dropped by 10 percent.*
- *The employee absence rate has increased by 20 percent.*

In a meeting with the board of directors, some of the senior managers expressed the view that Six Sigma works only in the Japanese automotive industry. Based on this information, the board decides to cancel the Six Sigma program.

1. Was scientific evidence consulted
2. Was organizational evidence considered?
3. Was professional experience and judgment taken into account?
4. Were stakeholder values and concerns considered?

Why do we have to critically appraise evidence?

Evidence is never perfect and can be misleading in many different ways. It may be that the evidence is overstated such that a seemingly strong claim turns out to be based on a single and not particularly reliable piece of information. A colleague's confident opinion regarding the effectiveness of a practice might turn out to be based on little more than an anecdote.

A long-standing way of doing things in an organization may actually never have been evaluated to see whether it worked optimally. All evidence should be critically appraised by carefully and systematically assessing its trustworthiness (certainty) and relevance. This appraisal is the critical-thinking part of evidence-based practice.

Although how a piece of evidence is evaluated can differ slightly depending on its source, critical appraisal always involves asking the same basic questions:

- Where and how is the evidence gathered?
- Is it the best available evidence?
- Is there enough evidence to reach a conclusion?
- Are there reasons why the evidence could be biased in a particular direction?

So, for example, if we are critically appraising a colleague's experiences with a particular problem, we may wonder how many times he or she has experienced that problem and whether the situations were comparable. If it turns out that our colleague has relatively little experience and in settings quite different from our own, we might judge this evidence to be irrelevant or weak. Similar questions need to be asked about organizational evidence such as sales figures, error rates, and cash flow. How were these figures calculated? Are they accurate? Are they reliable? In the case of scientific evidence, we would ask questions about how the study was designed. How were the data collected? How was the outcome measured? To what extent are alternative explanations for the outcome found possible? Evidence-based practice is about using the best available evidence. Judging the quality and strength of the evidence we have through the process of critical appraisal enables us to discern and identify the best available evidence.

Did I get this 1.4

How trustworthy would you consider the following evidence (very high, high, moderate, low, very low)?

1. Findings from a survey regarding the effect of financial incentives on employee performance conducted by an insurance company and published in a press release
2. Findings from a study regarding the effect of financial incentives on employee performance conducted by researchers from a university and published in the scholarly Journal of Organizational Behavior. The study compares the average performance of employees from 20 organizations that use financial incentives with 20 organizations that don't use financial incentives. Performance was measured both before and after the implementation.
3. The personal opinion of a senior manager regarding the effect of financial incentives on employee performance
4. The outcome of a survey among nurses in a local hospital that assessed their feelings toward the use of financial incentives to improve performance. The survey was completed by 90 percent of the nurses; all responded that they considered the use of financial incentives to improve performance to be unethical.

Explain the above rating as to how trustworthy you consider the evidence.

Why focus on the best available evidence?

In almost any situation, it is possible to gather different types of evidence from different sources. In many situations, it may also be possible to gather a large quantity of evidence. But which evidence should we pay most attention to, and why? A fundamental principle of evidence-based practice is that the quality of our decisions is likely to improve the more use we make of trustworthy evidence – in other words, ***the best available*** evidence.

This principle is apparent in everyday decision-making, whether it's buying someone a birthday present or wondering where to go out for dinner. In most cases, we actively seek information from multiple sources, such as our partner's opinion, the experiences of friends, or the comments of a local food critic. Sometimes, this information is so weak that it is hardly convincing at all; other times, the information is so strong that no one doubts its correctness. It is therefore important to be able, through critical appraisal, to determine what evidence is the best – the strongest, most trustworthy – evidence. For instance, when planning a vacation to the Emerald Isle, the most trustworthy evidence on which area of Ireland has the least chance of rain in early August will obviously come from statistics on the average rainfall per month, not from the personal experience of a colleague who visited the country only once.

Learn by doing 1.8

Trustworthiness of Evidence Sources

In this activity, consider the scenario and the trustworthiness of the four evidence sources.

Scenario: *You have recently been appointed as director for a medium-sized business. In your first months working there, it becomes clear to you that the company is not organized in the best way. Overhead is too high, the reward system is very old, and profit margins have halved over the last two years. The CEO wants this situation to change and is wondering whether the introduction of the Meyer-Whitney model could improve the financial performance of the company. Since you have never heard of this model, you decide to consult evidence from a number of sources before you give your advice to the CEO.*

The four evidence sources are as follows:

1. A recent scientific study published in an academic journal, which shows that the Meyer-Whitney model has no effect on the financial performance of an organization. The study compares 20 organizations that have implemented the Meyer-Whitney model with 20 organizations that have not implemented it. Measuring the financial performance before and after implementation showed that there was no difference between the two groups.
2. A multiple case study published in a popular management magazine, that appears to show that the financial performance of three British organizations showed a major improvement within a year after the Meyer-Whitney model was introduced.
3. In a national newspaper, an article was recently published in which the famous American CEO of a large multinational company talked about his experiences with the Meyer-Whitney model. In the article, he says that since the introduction of this model, the stock market value of the company has increased by 20 percent.

4. You contact a senior consultant at a well-known consulting firm. This consultant tells you that he does not think the Meyer-Whitney model has an effect on the financial performance of an organization. He advises you not to introduce the model.

Please read the scenario and the four evidence sources above and rank the evidence according to its trustworthiness: 1 = the strongest / most trustworthy evidence to 4 = the weakest / least trustworthy. Use each number only once.

As is illustrated in the preceding activity, exactly the same is true for management decisions. When making a decision about whether to use a quality management method such as Six Sigma to reduce medical errors in a British university hospital, information based on the findings from a study of 150 European university hospitals in which medical errors were measured before and after the introduction of Six Sigma is stronger (more trustworthy) than the professional experience of a colleague who works at a small private hospital in Sydney, Australia.

However, such a study may never have been done. Instead, the best ***available*** evidence could be case studies of just one or two hospitals. For some decisions, there may be no scientific or organizational evidence at all. Consequently, we may have no option but to make a decision based on the professional experience of colleagues or to pilot test some different approaches and see for ourselves what might work best. Given the principles of evidence-based practice, even if we rely on the experience of a colleague, using this limited-quality evidence can still lead to a better decision than not using it, as long as we are aware of its limitations when we act on it.

Did I get this 1.5

In the following activity, consider the scenario, then read the four evidence sources and rank the evidence based on its trustworthiness.

Scenario: *You are working as a manager at a large Italian IT firm. The productivity of the engineers is far below average for the sector. The board of directors wants to take action to improve this situation. The financial director suggests introducing a performance-related pay model, which would give the workers a financial incentive to carry out more work. Since you are not sure what causes the low productivity of the engineers, you feel it will be useful to consult evidence from a number of sources before you decide to implement a performance-related pay model.*

The four evidence sources are as follows:

1. The advice of a senior consultant at a well-known consulting firm, who tells you that, on the basis of hi 15 years of experience with a large number of IT firms in the United States, the most likely cause for lo productivity is lack of leadership. He therefore advises you not to introduce the performance-related pay model and instead to provide leadership training for all supervisors and senior managers.
2. The testimony of the chief operating officer, who tells you that the organizational data show that the engineers' productivity varies widely per team. In fact, analysis shows that teams with a large number of senior and experienced engineers are twice as productive as teams with a large number of young and inexperienced engineers. She therefore advises you to invest in the training and support of young and inexperienced engineers and to arrange a more balanced distribution between the teams.
3. A case study shown to you by the financial director; this case study, published in an academic journal, indicates that the productivity of Chinese mineworkers of four state-owned companies improved within a year after a performance-related pay model was introduced. The result of this case study is based on interviews with 50 supervisors. The financial director therefore advises you to introduce the performance-related pay model.

4. The testimony of the HR director, who tells you that she thinks the most likely cause for low productivity is lack of teamwork: it is well known that engineers in general lack social skills and consequently don't share among themselves the task-relevant knowledge that is necessary to solve practical problems and improve performance. She therefore advises you not to introduce the performance-related pay model and instead to invest in team building.

Please read the scenario and the four evidence sources above and rank the evidence according to its trustworthiness: 1 = the strongest / most trustworthy evidence to 4 = the weakest / least trustworthy. Use each number only once.

5. Now let's assume that in the preceding scenario, the only evidence available is a multiple case study published in a popular management magazine. Based on this study, the CEO decides to take his chances and implement the Meyer-Whitney model.

Would you consider this to be an evidence-based decision?

- Yes, because the CEO has consulted evidence from the scientific literature.
- No, because the case study is not the best available evidence.
- No, because a case study with no premeasure or comparison, published in a popular magazine, should be regarded as low-quality evidence.
- Yes, because the multiple case study is the best available evidence.

Common misconceptions of evidence-based practice

Misconceptions about evidence-based practice are a major barrier to its uptake and use. For this reason, it is important that misconceptions be challenged and corrected. In most cases, they reflect a narrow or limited understanding of the principles of evidence-based management.

Learn by doing 1.9

Based on what you have learned so far, are the following six statements true or false?

1. "Evidence-based management disregards managers' professional experience.".

Our answer:

This misconception directly contradicts our definition of evidence-based management: that decisions should be made through the conscientious, explicit, and judicious use of evidence from multiple sources, including experiential evidence. Evidence-based management does not mean any one source of evidence is more valid than another. Even experiential evidence – the professional experience and judgment of managers – can be an important source of evidence if it is appraised to be trustworthy and relevant. Experiential evidence is essential in appropriately interpreting and using evidence from other sources. If we are trying to identify effective ways to share information with colleagues, scientific and organizational evidence may be informative, but experiential evidence is needed to help figu e out what practices make good sense if we are working with professionally trained colleagues or relatively low-skilled workers. Similarly, scientific evidence can help us to understand the extent to which our experiential evidence is trustworthy. Research indicates that years of experience in a technical specialty can lead to considerable expertise and tacit knowledge. On the other hand, an individual holding a series of unrelated jobs over the same number of years may have less trustworthy, less reliable expertise. That's why evidence-based management is about using evidence from multiple sources rather than merely relying on only one.

2. "Evidence-based management is all about numbers and statistics."

Our answer:

Evidence-based management is about seeking and using the best available evidence from multiple sources. It's not all about numbers and quantitative data, though many management decisions involve numbers and figu es of some sort. You do not need to become a statistician to make an evidence-based decision, but it does help to have an understanding of some basic statistical concepts that are useful to critically evaluate some types of evidence. The principles behind such concepts as sample size, statistical versus practical significance, confidence intervals, and effect sizes can be understood without any math. Evidence-based management is not about doing statistics, but statistical thinking is an important element.

3. "Managers need to make decisions fast and don't have time to take an evidence-based approach."

Our answer:

Sometimes, taking an evidence-based approach is about taking a moment to reflect on how well the evidence you have can be trusted. More often, it is about preparing yourself to make key decisions well – by identifying the best available evidence you need, preferably before you need it. It is true that some organizational decisions must be made quickly, but even split-second decisions require trustworthy evidence. Making a fast and sound decision about when to evacuate a leaking nuclear power plant or how to make an emergency landing requires up-to-date knowledge of emergency procedures and reliable instruments providing trustworthy evidence about radiation levels or altitude. Likewise, when important decisions really must be made quickly, an evidence-based manager anticipates the kinds of evidence quality decisions require. However, the need to make really fast management decisions in many cases is the exception rather than the rule. Most decisions take place over much longer time periods – sometimes weeks or even months – and often require the consideration of legal, financial, strategic, and other organizational issues, which by necessity takes time. This provides plenty of opportunities to collect and critically evaluate evidence about the nature of the problem and, if there is a problem, the decision most likely to produce the desired outcome. For evidence-based management, time is not a deal breaker.

4. "Every organization is unique, so the usefulness of scientific evidence is limited."

Our answer:

One objection some people have to using research evidence is the belief that every organization is unique, suggesting that research findings simply will not apply. Although organizations do di fer, they also tend to face similar issues, sometimes repeatedly, and often respond to them in similar ways. Peter Drucker, a seminal management thinker, was perhaps the first to assert that most management issues are "repetitions of familiar problems cloaked in the guise of uniqueness."[14]

It is commonplace for organizations to have myths and stories about their own uniqueness[15] – it is not that all organizations are alike, nor is it that each is unique: the reality lies somewhere in between. As an evidence-based manager, you need to be flexible enough to take your organization s similar-yet-different qualities into account. For instance, an evidence-based manager may use individual financial incentives for sales agents but reward highly skilled knowledge workers with more opportunities for development or personally interesting projects, knowing that financial incentives tend to lower performance for knowledge workers.[16, 17]

5. "If you don't have high-quality evidence, you can't do anything."

Our answer:

Sometimes, there is very little or no quality evidence available. This may be the case, for example, with a new management practice or the implementation of a new HR model. In some areas, the organizational context changes rapidly, which can limit the relevance and applicability of scientific and experiential evidence derived in a context different from today's. In those cases, the evidence-based practitioner has no other option but to work with the limited evidence at hand and supplement it through learning by doing. This means pilot testing and treating any course of action as a prototype: systematically assessing the outcome of the decisions we take through a process of constant experimentation, punctuated by critical reflection about which things work and which things do not [18, 19]

6. "Good-quality evidence gives you the answer to the problem."

Our answer:

Evidence is not answers. It does not speak for itself. To make sense of evidence, we need an understanding of the context and a critical mindset. You might take a test and find out you scored 10 points. But if you don't know the average or total possible score, it's hard to figu e out if you did well. At the same time, you might want to know what doing well on the test actually means. Does it indicate or predict anything important to you and in your context? And why? Simply knowing you scored 10 points on the test is meaningless data without this additional information. At the same time, evidence is never conclusive. It doesn't prove things, which means no piece of evidence can be viewed as a universal or timeless truth. In fact, in most cases, evidence comes with a large degree of uncertainty. Evidence-based practitioners therefore make decisions not based on conclusive, solid, up-to-date information but on probabilities, indications, and tentative conclusions. Evidence doesn't tell you what to decide, but it does help you to make a better decision.

What is the evidence for Evidence-based Management?

People sometimes ask whether there is evidence that an evidence-based management approach is more effective than the way typical managers already make decisions. This is, of course, a very important question. To measure the effect of evidence-based management would require (1) an evaluation of a large number of situations and contexts where evidence-based practice was applied and (2) the measurement of a wide range of outcomes, preferably by means of a double-blind, randomized controlled study. Such a study might be too difficult to actually carry out. However, there is plenty of evidence that suggests that taking an evidence-based approach to decisions is likely to increase the effectiveness of those decisions.

We noted that the human mind is susceptible to systematic errors – we have cognitive limits and are prone to biases that impair the quality of the decisions we make. The fundamental questions to ask include, How can we make decisions without falling prey to our biases? Are there decision practices or processes that can improve decision quality? Fortunately, a large number of studies ranging from systematic reviews to qualitative research indicate that

- Forecasts or risk assessments based on the aggregated (averaged) professional experience of many people are more accurate than forecasts based on one person's personal experience (provided that the forecasts are made independently before being combined). [20] [21] [22] [23] [24]
- Professional judgments based on hard data or statistical models are more accurate than judgments based on individual experience. [25] [26] [27]
- Knowledge derived from scientific evidence is more accurate than the opinions of experts. [28]
- A decision based on the combination of critically appraised experiential, organizational, and scientific evidence yields better outcomes than a decision based on a single source of evidence. [29] [30]
- Evaluating the outcome of a decision has been found to improve both organizational learning and performance, especially in novel and non-routine situations. [31] [32]

We started this module by explaining what evidence-based practice is about: making decisions through the conscientious, explicit, and judicious use of the best available evidence from multiple sources. By using and critically appraising evidence from multiple sources, you increase the likelihood of making an effective decision.

We discussed why we need evidence-based practice. Most managers prefer to make decisions based solely on personal experience. However, personal judgment alone is not a reliable source of evidence because it is prone to cognitive biases and thinking errors. In addition, managers and consultants are often unaware of the current scientific evidence available: large discrepancies seem to exist between what managers and consultants think is effective and what the current scientific research shows. As a result, billions of dollars are spent on management practices that are ineffective or even harmful to organizations.

We then discussed what counts as evidence. When we say *evidence*, we basically mean information. It may come from scientific research, from the organization itself, or from the professional experience of managers. Even evidence regarding the values and concerns of stakeholders (e.g., employees who will be affected by the outcome of the decision) may be important to consider.

We also explained that evidence is never perfect. We must always critically appraise the trustworthiness of the evidence, regardless whether it is evidence from experience or evidence from scientific research. We can do that by asking how the evidence is gathered, if it could be biased in a particular direction, and if it is the *best available evidence*.

Sometimes, the best available evidence is hardly convincing at all; other times, the evidence is so strong that no one doubts its correctness. In fact, in some situations, there is little or no quality evidence available. In those cases, we have no choice but to work with the limited evidence at hand and supplement it through learning by doing. This means pilot testing and systematically assessing the outcome of the decisions we make.

Finally, you have learned that evidence is not answers. It does not speak for itself. In fact, in most cases, evidence comes with a large degree of uncertainty. Evidence-based practitioners therefore make decisions based not on conclusive, solid, up-to-date information but on probabilities, indications, and tentative conclusions.

The most important learning point is that evidence-based practice starts with a critical mindset. It means questioning assumptions, particularly when someone (including ourselves) asserts some belief as a fact. So, from now on, always ask, *What's the evidence for that? How trustworthy is this evidence? and Is this the best available evidence?*

Podcast: Principles of Evidence Based Management

In this first podcast host Karen Plum discusses with Eric Barends, Managing Director of the Center for Evidence Based Management (CEBMa), Denise Rousseau, Professor of Organizational Behavior and Public Policy at Carnegie Mellon University, and Rob Briner, Professor of Organizational Psychology at Queen Mary University of London the basic principles of evidence-based management.

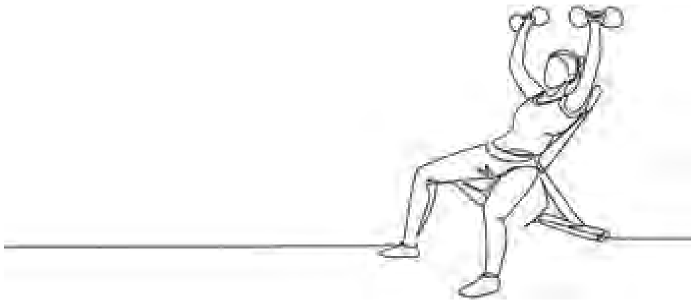
What is evidence-based management and why do we need it? Because everyone uses some kind of evidence when making decisions, right? But only one source, or many? And are those sources assessed for their reliability and trustworthiness? Are we actively trying to identify the biases that so often lead us down the wrong path?

And really, who has time to take an evidence-based approach, when organisations just want to “get things done”?

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 1.1: How do you typically make decisions?

Think about how you make decisions as a consumer (e.g., buying a new laptop, choosing a restaurant, booking a hotel). Take two decisions: one you are satisfied with and one you a e not as satisfied with. rite out your answers in two columns to compare your answers to the following questions.

1. *How did your decision-making process look like? (e.g., How did you start? How many alternatives did you consider?)*
2. *What evidence did you use to decide between alternatives?*
3. *Did you use product review sites such as Yelp, TripAdvisor, or Amazon reviews? Why / why not?*
4. *Overall, how trustworthy was the evidence you did use?*
5. *How long did it take before you made your final decision?*
6. *On comparing the two decisions, what reflections do you have on your decision processes?*

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

Exercise 1.2: How does your organization typically make decisions?

Download the "How evidence-based is your organization?" **self-assessment survey**.

Fill out the survey with two or more colleagues and submit the results (and possible questions & comments) according to the instructions of your professor/instructor.

Exercise 1.3: How does your organization typically make decisions? – Post Mortem

Think back to a management, business, or policy decision you have been involved in making (or have observed closely). This decision should be reasonably important for your organization, involving significant resources, several or many people, and with no 'easy' answer. Write your answers to the following three questions in detail.

1. *What exactly was the problem to be solved (or opportunity to be addressed)?*
2. *What evidence was available and from which sources?*
3. *Was any attempt made to explicitly evaluate its quality or trustworthiness?*
4. *What was the decision-making process? (Steps taken, alternatives considered, stakeholders involved, and who made the final decision?)*
5. *What could have been done to improve the 'evidence base' of the decision - what specific types of evidence would have been helpful?*

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Exercise 1.4: Self-guided book tour

We would like you to get a hold of three popular/best-selling management books. You can go the bookstore, check out your library (or a friend's) or go on-line. Pick three books that interest you. For each book, indicate what its 'evidence-base' appears to be. Write down your answers to the following questions:

1. *What types of evidence sources does the book cite? In particular, what proportion of sources appears to be*
   - *personal/anecdotal?*
   - *'best-practices' of other companies?*
   - *other business books or publications?*
   - *evidence from scientific research?*
2. *What is known about the author? (check Google or Google Scholar)*
3. *What is known about the proposed model/principles/insights? (what's the evidence base for its proposed model or insights?)*
4. *Which book appears to you to be the most useful? Why?*
5. *Does being evidence-based affect usefulness?*

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Suggestions for further reading

- *Managerial fads and fashions: The diffusion and rejection of innovations*, Abrahamson, AMR, 1991
- *Surprising but true: Half the decisions in organizations fail*, Nutt, AMP, 1999
- *What bandwagons bring: Effects of popular management techniques on corporate performance, reputation, and CEO pay*, Staw & Epstein, ASQ, 2000
- *Seven common misconceptions about human resource practices: Research findings versus practitioner beliefs*, Rynes et al, AMP, 2002
- *Evidence Based Management*, Pfeffer & Sutton, Harvard Business Review, 2006
- *Trust the Evidence, Not Your Instincts*, Pfeffer & Sutton, The New York Times, 2011
- *Evidence Based HR: Under The Microscope*, Katie Jacobs, HR Magazine, 2015
- *The Basics of Evidence-Based Practice*, Rob Briner, HR People & Strategy, 2018

References

1. Pfeffer, J., and Sutton, R. (September 3, 2011). “Trust the evidence, not your instincts.” *New York Times*.
2. Mintzberg, H. (1975). “The manager’s job: Folklore and fact.” *Harvard Business Review*. Volume 53.
3. Dawes, M., Summerskill, W., Glasziou, P., et al. (2005). “Sicily statement on evidence-based practice.” *BMC Medical Education*. Volume 5. Number 1.
4. Kahneman, D. (2011). *Thinking, fast and slow*. Penguin Group. London.
5. Clements, M. P. (2003). “An evaluation of the Survey of Professional Forecasters probability distribution of expected inflation and output g owth.” *Journal of Economic Literature*. Volume 41. Number 4.
6. Bazerman, M. H. (2009). *Judgment in managerial decision making*. Wiley. New York.
7. Simon, H. A. (1997). *Models of bounded rationality*. MIT Press. New York. Vol. 3.
8. Barnett-Hart, A. K. (2009). *The story of the CDO market meltdown: An empirical analysis*. Harvard University. Series M-RCBG associate working paper no. 4.
9. Loungani, P.(December 18, 2000). “The arcane art of predicting recessions.” *Financial Times*.
10. Tetlock, P. E. (2006). *Expert political judgment: How good is it? How can we know?* Princeton University Press. Princeton, NJ.
11. Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). “Improving patient care—systematic review: The relationship between clinical experience and quality of health care.” *Annals of Internal Medicine*. Volume 142. Number 4.
12. Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). “HR professionals’ beliefs about effective human resource practices: Correspondence between research and practice.” *Human Resource Management.* Volume 41. Number 2. 149-174.
13. Biography.com Editors (2014). *Chesley Sullenberger Biography*. The Biography.com website. https://www.biography.com/personality/chesley-sullenberger.
14. Lowenstein, R.(January 22, 2006). “When business has questions, Drucker still has answers.” *New York Times*.
15. Martin, J., Feldman, M. S., Hatch, M. J., & Sitkin, S. B. (1983). “The uniqueness paradox in organizational stories.” *Administrative Science Quarterly*. 438-453.
16. Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). “Large stakes and big mistakes.” *Review of Economic Studies*. Volume 76. Number 2.
17. Joseph, K., & Kalwani, M. U. (1998). “The role of bonus pay in salesforce compensation plans.” *Industrial Marketing Management*. Volume 27. Number 2.
18. Pfeffer, J., & Sutton, R. I. (2010). “Treat your organization as a prototype: The essence of evidence-based management.” *Design Management Review*. Volume 17. Number 3.
19. Weick, K. E, & Sutcliffe, K. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty*. New York: Wiley.
20. Silver, N. (2012). *The signal and the noise: Why so many predictions fail—but some don’t. London:* Penguin. pp. 286, 690.
21. Bauer, A., Eisenbeis, R. A., Waggoner, D. F., & Zha, T. (2003). “Forecast evaluation with cross-sectional data: The blue chip surveys.” *Economic Review*. Volume 88. Number 2. 17-32 Pages.
22. Servan-Schreiber, E., Wolfers, J., Pennock, D. M., & Galebach, B. (2004). “Prediction markets: Does money matter?.” *Electronic Markets*. Volume 14. Number 3. 243-251 Pages.

23. Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Chapter Combining Forecasts. New York: Kluwer Academic.

24. Yaniv, I., & Choshen-Hillel, S. (2011). "Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy." *Journal of Behavioral Decision Making*. Volume 25. Number 5. 427-434.

25. Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York: W. W. Norton.

26. Grove, W. M. (2005). "Clinical versus statistical prediction." *Journal of Clinical Psychology*. Volume 61. Number 10. 1233-1243.

27. Ayres, I. (2007). *Super crunchers*. New York: Bantam Books.

28. Antman, E. M., J Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). "A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts." *JAMA*. Volume 268. Number 2. 240-248.

29. McNees, S. K. (1990). "The role of judgment in macroeconomic forecasting accuracy." *International Journal of Forecasting*. Volume 6. Number 3. 287-299.

30. Tetlock, P. E. (2006). *Expert political judgement*. Princeton, NJ: Princeton University Press.

31. Anseel, F., Lievens, F., & Schollaert, E. (2009). "Reflection as a strategy to enhance task performance after feedback." *Organizational Behavior and Human Decision Processes*. Volume 110. Number 1. 23.

32. Ellis, S., & Davidi, I. (2005). "After-event reviews: Drawing lessons from successful and failed experience." *Journal of Applied Psychology*. Volume 90. Number 5. 857-871.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 1.1

Our answer: C and D

Learn by doing 1.2

1: Yes > The experience of employees with flexible working arrangements can be conside ed as evidence. Evidence such as employees' experience can provide important information about whether a new way of working or a new procedure is feasible or effective in a particular context.

2: No > It is unclear what common knowledge means: findings f om an employee survey, the collective personal experience of managers? When no supporting evidence is provided, assertions based on common knowledge should be considered a belief or personal opinion.

3: Yes > Even one person's professional experience counts as evidence.

Did I get this? 1.1

Yes > It is unclear what the evidence is for the CEO's assertion: professional experience, findings f om scientific esearch? As long as you can ask, "How do you know?" or "What is the evidence?" it is likely that the information presented is not evidence but an assumption, a personal opinion, or a belief. The fact that the CEO has a degree from a prestigious university does not mean his or her assertion is evidence.

Learn by doing 1.3

Our answer: Research suggests that most management practitioners base their decisions on intuition and experience. In fact, practitioners seldom consult the findings of scientific esearch. The most important source to which practitioners turn when faced with a management problem seems to be their colleagues or the insights from "experts."

Learn by doing 1.4

Our answer: A, C and D

Learn by doing 1.5

1: Evidence from the scientific literatu e refers to the findings f om empirical studies published in academic journals.

2: Evidence from the organization refers to data, facts, and figu es that are generated by the organization itself and measured repeatedly over time. It can be hard numbers such as financial data, tu nover rates, or client satisfaction. It can also include soft elements, such as organizational characteristics (e.g., cultural norms or structural information) as well as the outcomes of exit interviews.

Learn by doing 1.6

1: Experiential evidence refers to evidence from the professional experience, knowledge, skill, and expertise of practitioners. This type of evidence is sometimes referred to as *tacit knowledge*.

2: Stakeholder evidence refers to what individuals or groups affected by a decision believe to be important. Stakeholder evidence is relevant to how an individual or group might react to a decision's possible consequences.

Did I get this? 1.2

1: Yes > The public opinion or other expressions of stakeholder interests and concerns are important stakeholder evidence that may influence the outcome of decision-making

2: Yes > The outcome of employee satisfaction surveys can be considered as organizational evidence: evidence that is generated by the organization itself and is repeatedly measured over time.

3: Yes > Employees are important stakeholders. Their fears and concerns, regardless whether they are based on facts or rumors, can be considered as stakeholder evidence. Although stakeholders' concerns can be highly subjective and even sometimes seemingly irrational, they provide important information about how a proposed decision might be received.

Learn by doing 1.7

1: No > No scientific evidence was consulted. Based on the outcome of the meeting with the employees, the scientific literatu e could have been consulted to find out if esearch findings confirm a positiv effect of trust in management on employee satisfaction.

2: Yes > The outcome of the satisfaction survey can be considered as organizational evidence: evidence that is generated by the organization itself and is repeatedly measured over time.

3: No > The employees were asked to give their 'opinion', which is not the same as professional experience and judgment.

4: Yes > The outcome of the meeting with the employees can be considered as stakeholder evidence.

Did I get this? 1.3

1: No > No scientific evidence was consulted. The boa d could have asked these questions: What does scientific esearch suggest to be the major cause of declining productivity or increasing absence rates? Given the target group and the context involved, what are the main factors determining the success or failure of Six Sigma?

2: Yes/No > Partially correct. The number of problem analyses and improvement measures, the number of breakdowns and short stops, the average productivity per worker, and the employee absence rate all can be regarded as organizational evidence. However, it could be argued that important organizational evidence is missing: Is there a correlation between absence rate, productivity, short stops, and the implementation of Six Sigma? Is there a trend? What is the average in the sector?

3: No > The views of some senior managers cannot be regarded as experiential evidence. When no supporting evidence is provided, views should be considered a belief or personal opinion. The board could have asked these questions: Have they seen this scenario before? What are their experiences regarding declining productivity in general? What do they think are the causes in this particular case? How relevant and applicable is their experience?

4: No > In this case no stakeholder evidence was consulted. Relevant questions include, How do employees feel about Six Sigma? Do they see downsides or unintended negative consequences? How practical or workable do those responsible for implementing Six Sigma feel? What do they think are the causes of the declining productivity?

Did I get this? 1.4

1: Our answer: Findings from a survey conducted by an insurance company and published in a nonacademic journal such as a press release, newspaper, or magazine cannot be considered as (highly) trustworthy. The evidence could be stronger / more trustworthy if it came from a survey conducted by scholarly researchers who strive to be objective and if it were published in a journal that maintains a high standard of quality, accuracy, and academic integrity. Research conducted by organizations and published in press releases, newspapers, or magazines could be biased or influenced by the desi e to obtain certain findings

2: Findings from a study that measures the effect of financial incentives on employee performance both before and after the implementation and compares the outcome with other organizations not using financial incentives can be egarded as (highly) trustworthy evidence. In addition, findings f om a study conducted by independent researchers and published in an academic journal such as the *Journal of Organizational Behavior* are more trustworthy than findings f om a study published in a popular magazine.

3: If no supporting evidence is given, the personal opinion of a senior manager is not very trustworthy: it is a belief because it reflects a subjective point of vie . However, senior managers are important stakeholders in the decision making process, so from an evidence-based perspective, it could be important to take their personal opinions as stakeholders into account.

4: The unanimous outcome of a survey with a response rate of 90 percent can be regarded as (very) trustworthy evidence: in this case, it is a strong indicator that financial incentives would be unacceptable to these nurses.

Learn by doing 1.8

1: 3 > This is the professional experience of only one person, published in a newspaper. In addition, the outcome (stock market value) is not very relevant. We would argue it is less strong/trustworthy than the scientific study or the case study, though mo e trustworthy than the personal opinion of a consultant.

2: 4 > No evidence is provided, so the consultant's advice should be considered a personal opinion or belief. This makes it less strong/trustworthy than the scientific study, the case study, and the CEO s personal experience.

3: 1 > This is a controlled study with a before-and-after measurement, published in an academic journal, so this is a very strong/trustworthy source of evidence.

4: 2 > This is a case study of only three organizations, with no premeasure or comparison, so it is less strong/trustworthy than a controlled study of 20 + 20 organizations. However, it is more trustworthy than the professional expertise of one person or the personal opinion of a consultant.

Did I get this? 1.5

1: 4 > No evidence is provided, so the HR director's testimony should be considered a personal opinion or belief.

2: 2 > This evidence represents a scientific study published in an academic jou nal. However, the case study is based on just a single organization and uses no premeasures or controls, making bias likely in what is reported. Also, note that the study population and context are quite different from the focus of this decision. This makes it less reliable and applicable than objective and reliable data from the organization itself.

3: 1 > This evidence is based on solid, objective, and reliable data from the organization itself in which a direct comparison between teams with experience and teams of less experienced engineers was made. This makes it stronger / more trustworthy than the other sources of evidence.

4: 3 > This evidence represents the professional experience of just a single person. It is not clear what measures of productivity and teamwork were used. Thus, the experience may be based on a general impression rather than reliable or valid data; it also is based in a different context (a US instead of Italian firm)

5: Yes > Although this is just a case study with no premeasure or comparison, published in a popular magazine, it is still the best available evidence.

Learn by doing 1.9

Answers are provided in the module.

CEBMa
center for
Evidence-Based Management

Open Learning Initiative

MODULE 2 | ASK Critical Questions about Problems and Solutions

Learning objectives:

- Identify claims, assumptions, and hypotheses regarding a practical issue.
- Determine what the (assumed) problem is for which the evidence should be consulted.
- Formulate questions to determine whether the (assumed) problem is supported by evidence from multiple sources.
- Determine what the preferred solution is for which the evidence should be consulted.
- Formulate questions to determine whether the preferred solution is supported by evidence from multiple sources.

Introduction

> *The important thing is to never stop questioning.*
>
> *– ALBERT EINSTEIN*

At the outset of a decision, it is critical to figure out the problem you need to solve or the opportunity or you are trying to address. A good start is to ask questions – lots of them. Asking questions kicks off the process of a deliberate search for evidence and understanding, gathering intelligence to get a full grasp of the need, the opportunity, or the crisis. Asking questions to identify uncertainties – and thus the need for evidence – is therefore the first step of evidence-based management. In fact, asking questions and framing diagnoses and problems properly may be the most important step in the evidence-based process.

Indeed, this first step of evidence-based management involves learning to ask the right questions, as illustrated by the American astrophysicist Neil deGrasse Tyson. [1]

Example

Imagine someone comes up to you and says, "I have these crystals. If you rub them together, it will heal all your illnesses. I'm happy to sell them to you for one hundred dollars." What would your response be? Would it be, "Oh, great. Wow! Here's the money"? Or would you say, "Oh, that's rubbish. That will never work"? Each of those responses is equally scientifically lazy. In the first case, you say it's definitely true; in the second case, you say that it could never be true. But neither response – neither acceptance nor rejection – requires thought. So, extreme gullibility and extreme skepticism are two equal ways of not having to think much at all. It's harder to ask good questions. For example, "Where did you get the crystals? What are the crystals made of? What kind of diseases do you say they cure? How do you know it works? By what mechanism does it work? What evidence do you have that it would work on me? Can you demonstrate?" By the time you've finished, the person will probably have walked away.

Dismissing something as nonsense or embracing something as the truth before you have fully considered the evidence, even when it is the result of critical thinking or logical reasoning, is risky. Often, when we assume we are thinking critically or logically – that is, when we think we have a sound reason for our judgment – our decision-making may be tainted by cognitive biases and systematic errors. You can read more about biases and errors in Module 4, APPRAISE Evidence from Practitioners.

A better approach is to ask questions to determine whether strong evidence exists to support a claim, hypothesis, or assumption regarding a problem or solution. Thus, asking questions is the first step of evidence-based management. Is the claim a person makes based on strong/trustworthy evidence? Evidence-based professionals try always to maintain an open mind and a healthy dose of skepticism. They consistently (respectfully) question the information they are given, whether it is from their superior, a consultant, or a highly esteemed professor. In this module, you will learn how to ask questions to identify uncertainties and the need for evidence.

Identifying underlying assumptions

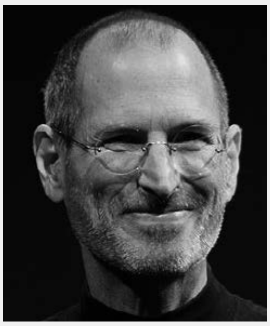
The main function of asking questions is to identify assumptions. An assumption is a claim, assertion, or hypothesis that we believe (or accept) to be true even though there is no evidence available (yet). In daily life and in the context of organizations, we make assumptions all the time. We assume that our car will be in the same spot we parked it in yesterday. We assume that our company did not burn down during the night. We assume that we will receive our paycheck at the end of the month.

Some assumptions turn out to be based on solid evidence, while others may have no supporting evidence, and some may even be false. In daily life, most of the assumptions we make are rather harmless and won't have serious consequences if they turn out to be incorrect. In the realm of management and organizations, however, an assumption underlying an important business decision that turns out to be based on fiction rather than on solid evidence can have a devastating impact. It may affect the company's business results and damage the working lives of employees. It is therefore important to identify assumptions underlying important managerial decisions and to check whether they are based on evidence.

A key problem with assumptions is that they are sometimes hidden. For example, consider the following assertion: "Teenagers nowadays spend hours sitting behind their computer. Consequently, their school performance will suffer." On close inspection, you will notice there are two hidden assumptions underlying this assertion. "Teenagers nowadays spend hours sitting behind their computer. *They don't use the computer to do their homework but to play games. Because they play games, they don't have time to do their homework.* Consequently, their school performance will suffer."

The same applies to the following claim: "Most organizations with an HR department have a low absenteeism rate, so all organizations should have an HR department." Here, the hidden assumption is that an organization's low absenteeism rate is attributable to the HR department – excluding all other possible explanations. The question to ask regarding such claims is whether the (hidden) assumptions are correct – that is, whether they are supported by strong/trustworthy evidence. For this reason, it is important not only that you identify explicit assumptions but also check for assumptions that are hidden.

Learn by doing 2.1



Read the following claims and identify the (hidden) underlying assumption for each claim. Select the *best* answer.

1: "Apple is a very successful organization; therefore, Steve Jobs must have been a great leader."

A. Successful companies have great leaders.

B. Steve Jobs was a great leader.

C. Steve Jobs is responsible for Apple's success.

D. A company's success is a direct result of its leadership.

2. *"To improve our company's productivity, we should increase employee satisfaction."*

A. A company's productivity is affected by the satisfaction of its employees.

B. The company's productivity is too low.

C. The company's employee satisfaction should be improved.

D. The company should spend more money on employee satisfaction.

3. *"At 3M, one of the most innovative companies ever, all employees can use up to 15 percent of their time to pursue new, innovative ideas, so we should introduce a similar rule."*

A. Our company is comparable to 3M.

B. Allowing employees to use up to 15 percent of their time to pursue new, innovative ideas will increase a company's innovativeness.

C. Our company should give employees more time to pursue new, innovative ideas.

D. What works for 3M will also work for other companies.

Did I get this 2.1

Read the claim below and identify the underlying (hidden) assumption. Select the best answer.

"Most organizations with a large HR department have a low absenteeism rate. Therefore, our company should increase its HR department

A. Most organizations with a high absenteeism rate have a small HR department.

B. Organizations with a high absenteeism rate should increase their HR department.

C. Absenteeism affected by the size of a company's HR department.

D. A low absenteeism can't be attributed to the presence of a large HR department.

As mentioned, people make assumptions all the time, and the same is true for managers, policymakers, and business leaders. Documents such as policy papers, project proposals, strategy documents, and change plans are often rife with both hidden and explicit assumptions. Obviously, it would not make sense – nor would it be feasible – to check the evidence for each assumption that is made, as most assumptions are rather trivial. However, the opposite is true for "critical" assumptions. When a critical assumption turns out to be false, the policy plan or decision may have severe negative consequences. Thus, the purpose of asking questions in the first stage of the evidence-based process is to

1. Identify critical assumptions, and
2. Check whether there is sufficient evidence to support these assumptions

Example

In February 2013, Marissa Mayer, CEO of Yahoo!, sent a memo to all her 12,000 employees stating that it is critical that all employees are present in their offices. “Some of the best decisions and insights come from hallway and cafeteria discussions, meeting new people, and impromptu team meetings. Speed and quality are often sacrificed when we work from home. We need to be one Yahoo!, and that starts with physically being together.” She therefore canceled all work-from-home arrangements and instructed all employees to work in the Yahoo offices. “Being a Yahoo isn’t just about your day-to-day job, it is about the interactions and experiences that are only possible in our offices.

While you may already have a strong opinion regarding the wisdom of the CEO’s decision to eliminate all work-from-home arrangements, the first step in the evidence-based process is to identify the most critical assumption(s) underlying this decision. In this case, there are critical assumptions that, if found to be incorrect (or overstated), may have a severe negative impact on the company.

Did I get this 2.2

In the example of Yahoo!, what would you consider to be the TWO most critical assumption(s) that should be checked to see whether there is sufficient evidence to support them?

A. Some of the best decisions and insights come from hallway and cafeteria discussions and meeting new people.

B. Speed and quality are often sacrificed when people work f om home.

C. To be one Yahoo!, employees need to be physically together.

D. True interactions and experiences are only possible in Yahoo!’s offices

E. Being a Yahoo isn’t just about employees’ day-to-day job.

If these two critical assumptions turn out to be incorrect, the CEO’s decision may have a negative impact not only on the job satisfaction and commitment of the employees in question but also on the creativity and performance of the company’s workforce as a whole. So, after identifying these statements as being critical assumptions, the obvious next question is, “What is the evidence for these claims?”

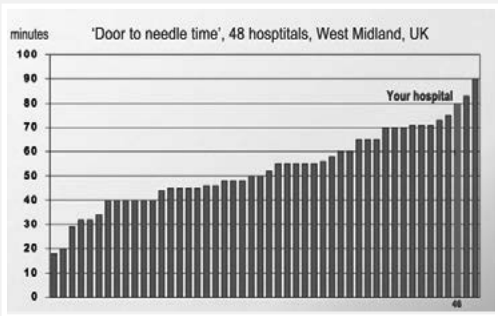
Starting Point: What is the problem to solve?

In evidence-based management, the starting point for asking questions is the assumed problem or opportunity rather than the preferred solution. Often, we dedicate a significant amount of time to determining exactly what the problem or opportunity might be. You may ask why a precise definition of the issue is important. Consider how crucial it is in the context of health care: there are strong indications that wrong diagnoses account for the most severe cases of patient harm. In fact, the leader of a recent study on misdiagnosis stated that "there's a lot more harm associated with diagnostic errors than we ever imagined. [2]

You may wonder if this is also true for management. When managers or business leaders decide to take action to address an assumed problem or opportunity, how accurate and reliable is their diagnosis? If the definition of the problem is incorrect, the problem cannot be addressed effectively even with an evidence-based approach. We therefore sympathize with this famous quote by Albert Einstein: "If I were given one hour to save the world, I would spend 59 minutes defining the problem and one minute solving it." In fact, our experience is that, in some cases, there is not even a problem to be solved.

For example, an organization may learn about a new, interesting solution (e.g., talent management), assume that the company somehow will benefit f om it, and thus decide to implement this new solution. However, implementing a solution when there is no evident problem or real opportunity makes little sense and can be a serious waste of time and resources. In the Yahoo! example, the preferred solution is clear (eliminating all work-from-home arrangements), but the assumed problem is less apparent. Is it poor performance, lack of creativity and innovation, or low product quality? And what is the evidence that these problems really exist? For this reason, an evidence-based approach always starts with the question, What is the problem you are trying to solve, and what is the evidence for this problem?

Learn by doing 2.2



In the United Kingdom, every year, 100,000 people have a stroke – an interruption of the blood supply to a part of the brain. Within minutes, brain cells begin to die. For this reason, a stroke is a medical emergency, and prompt treatment is crucial. In most cases, the best treatment is thrombolytic therapy, which is intravenous administration of a drug that dissolves blood clots. Obviously, the time between a patient with symptoms of a stroke enters the hospital and the start of the thrombolytic therapy should be as short as possible. This time is commonly referred to as the door to needle time (DNT). The standard in the UK is a maximum DNT of 55 minutes.

Imagine you are in charge of a hospital in West Midland, UK. A recent study by the National Health Service shows that your hospital has a mean DNT of 80 minutes, which is far above the standard. Obviously, this is a problem. But what is the problem behind the problem? What is the reason your hospital's DNT is far above the standard?

List all possible causes (underlying problems) for your hospital's long DNT. Think out of the box!

As you can see in the *Learn By Doing* exercise, there are many (more or less plausible) underlying causes that may have created the problem, and often these causes are very different in nature. For example, in our list of possible causes, number 1 concerns an educational problem, number 2 a payment problem, number 3 an IT problem, number 4 a procurement problem, number 5 a recruitment problem, and number 6 an architectural problem. Later in this module, we discuss how to differentiate problem symptoms from the underlying causes. However, first we discuss a conceptual tool that will help you to determine the organizational context

PICOC

When asking questions, it is important to make explicit the professional or organizational context that you should take into account. This is especially important when you direct questions to people outside the organization or consult external sources such as the research literature. For example, a question such as “Does team-building work?” may make sense in the context of your organization, but when consulting external sources, the question is obviously too vague. After all, you may be interested to know whether team-building in the form of an outdoor survival game improves the performance of a team of newly hired call center workers. Or whether team-building in the form of working with a coach may improve collaboration among a group of surgeons who have a poor relationship with each other. To make your question more context-specific, it helps to formulate what’s called a PICOC (Population, Intervention, Comparison, Outcome, Context). A PICOC is a conceptual tool designed to help you find evidence that takes into account your professional context, as explained in the following table.

Population	*Who?*	Type of employees
Intervention	*What or how?*	Management technique/method, factor, or independent variable
Comparison	*Compared to what?*	Alternative intervention, factor, or variable
Outcome	*What are you trying to accomplish, improve, or change?*	Objective, purpose, goal, or dependent variable
Context	*In what kind of organization or circumstances?*	Type of organization, sector, relevant contextual factors

In the Yahoo! example we used earlier, we could formulate the PICOC as follows:

P: IT workers, knowledge workers

I: Work-from-home arrangements

C: Traditional work arrangements

O: Task performance, creative performance

C: A multinational technology company

The underlying thought is that all five elements a e relevant to your questions and that each change in the **P**, **I**, **C**, **O**, or **C** may lead to a different answer. Thus, a general question such as whether work-from-home arrangements effectively yields answers of limited practical value because only the I (work-from-home arrangements) is addressed in the question, without considering

P: The effect may be different for blue-collar workers than for knowledge workers.

C: The effect may be different for agile working than for traditional working.

O: The effect on performance is possibly different from the effect on employee satisfaction.

C: The effect may be different for a tech company than for an academic hospital.

In short, your PICOC will help you to determine whether evidence from external sources (e.g., the findings of a scientific study) will be generalizable and applicable to your organizational context. For this reason, defining your PICOC is an important element of evidence-based management.

Learn by doing 2.3

Read the following scenario and determine the PICOC terms:

Many US banking firms use some sort of financial incentive plan to motivate their bank tellers by tying financial compensation and bonuses to their task performance. A lecturer at a community college in Canada wonders whether this practice would have a larger impact on students' academic achievement than does the current practice of setting individual performance goals.

P: What is the population of interest?

- Students
- Lecturers
- Bank tellers
- Canadian college students

I: What is the intervention (management technique, method) of interest?

- Academic achievement
- Financial incentives
- Performance goals
- Individual performance goals

C: What is the comparison (alternative method)?

- Intrinsic motivation
- Setting individual performance goals
- A financial incentives plan
- Goal setting

O: What is the outcome of interest?

- Student satisfaction
- Student grades
- Improved task performance
- Academic achievement

C: What is the context (what kind of organization or circumstances)?

- Any organization
- A Canadian community college
- An educational institution
- A financial organization

Did I get this 2.3

Read the following scenario and determine the PICOC terms:

According to the annual employee satisfaction survey of a Brazilian insurance company, one in three full-time employees feels that maintaining a healthy work–life balance has become more difficult in the last five years. In addition, the findings show that more than 80 percent of employees who fall within the millennials age bracket (individuals born between 1980 and 2000) are having an increasingly hard time maintaining a healthy work–life balance. Confronted with this disturbing outcome, the company's HR director suggests introducing flexible working hours for this particular group of employees. The director's assumption is that flexible working hours enables employees to determine the best way to allocate their time, attention, and energy resources, thereby increasing the likelihood that they can maintain a better work–life balance.

P: What is the population of interest?

- Employees who are having a hard time maintaining a healthy work–life balance
- Full-time employees
- Employees of a Brazilian insurance company
- Millennials

C: What is the comparison of interest?

- Doing nothing – business as usual
- Flexible working hours
- Maintaining a healthy work–life balance
- Employees who are not considered millennials

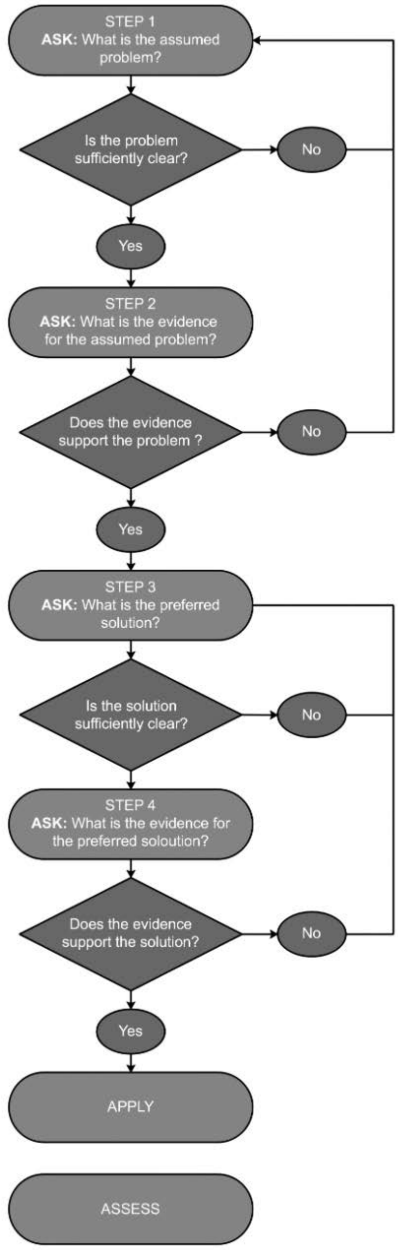
O: What is the outcome of interest?

- Flexible working hours
- A better work–life balance
- Increased employee satisfaction among millennials
- Increased employee satisfaction

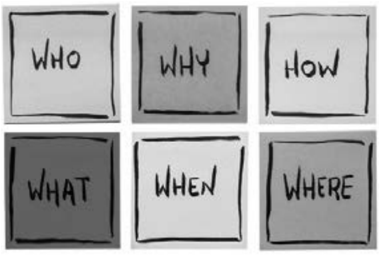
Overview of Questions

In the next sections, we provide an overview of questions you can ask to determine whether there is evidence to support (or contradict) an assumed problem or opportunity or a preferred solution. In most organizations, decisions are made not by one person but rather by a group of people, such as a board of directors, a committee, or a project team. In those situations, you would ideally take the role as “chief evidence officer,” monitoring and safeguarding the quality of the decision-making process by asking the questions shown in the below figure.

If you are the sole decision maker, however, it will be hard to question yourself. In that case, you should ask another person to take the role of critical enquirer to identify the underlying assumptions – and thus the need for evidence – in your reasoning. As we go forward, we will talk about both problems and opportunities under the label of “problems to be solved.” The figure below shows an overview of the process you would typically apply.



Step 1: What is the (assumed) problem to be solved?



As explained in the previous section, our first question is, What is the problem to be solved? In most organizations, an assumed problem is often composed of several underlying assumptions. Thus, “chunking,” or breaking the problem down – decomposing it into smaller, more specific, problems – is often useful, particularly when you find the problem overwhelming or daunting. For instance, in the Yahoo! example described earlier, the problem may be that the company’s performance is below the average in the sector, but a smaller, more specific underlying problem is that speed and quality are sacrificed when people work f om home. If the organization struggles with multiple problems at the same time, it is advisable to start with the most serious and urgent one (see question 3). When you have a satisfactory answer to what the most important problem is, you can ask five follow-up questions

1. How clearly defined is the problem?

Having a clear description of the assumed problem is the best first step to solving it because if you do not clearly define the problem, you probably cannot solve it. A good problem description entails at least five elements: What? Who? Where? When? Why?

2. Is it clear what the organizational consequences of the problem are?

A problem is only a problem when it has (potential) organizational consequences. For example, a low level of job satisfaction may be a problem only if it negatively affects a company’s business objectives (e.g., patient outcomes, innovation, or net profit margin) or the interests of its stakeholders (e.g., employee well-being, share price, or demands for community services). It is therefore critical that you clearly state how the problem affects or might affect important aims and outcomes.

3. Is it clear how serious and urgent the problem is?

Note that there is a difference between ***serious*** and ***urgent***. Serious problems have a potentially sizable impact on the organization’s aim and outcomes. Urgent problems are time-sensitive, demanding immediate attention, but their consequences may not necessarily impact significant aims and outcomes. Often, what we assume to be serious (important) problems are really just urgent ones, and not very serious or consequential at all. When we know which problems are both serious and urgent, we can move from firefighting – solving urgent but unimportant problems – to solving “real” problems.

4. Is it clear what the major cause(s) of the problem could be?

The key to a good definition of a problem is ensuring that you deal with the real problem – not its symptoms. What we often refer to as ***causes*** may actually be symptoms or indicators of deeper root causes. For example, a low level of job satisfaction is not the cause of a high employee turnover rate but the symptom of an underlying cause, such as underpayment or limited career opportunities. Problem symptoms and problem causes can look very much alike. It is therefore important that you differentiate symptoms from causes, for example, by continually asking, “Why is this issue occurring?” to each explanation and to subsequent explanations until you identify the root cause.

5. Is it clear what the logic model is?

A logic model spells out the process by which an underlying cause leads to a problem and produces certain organizational consequences. It is a short narrative that explains why or when the problem occurs (cause) and how this leads to a particular outcome (effect). In the Yahoo! example, we might describe the logic model as follows: *People who work at home are often distracted by all kinds of domestic and family issues. > Thus the speed and quality of their work are often sacrificed. > Yahoo! has many employees with a work-from-home arrangement. > This negatively affects the performance of these employees. > The company as a whole therefore performs below the average in the sector.*

Learn by doing 2.4

Read the following scenario and determine whether the problem is sufficiently clear:

SBC is one of the world's largest investment banks. Trades made by the firm's investment bankers in the front office are confirmed by Market Operations, which verifies and logs the trades and then transfers the money associated with a trade across accounts. Recently, senior executives in Market Operations became concerned with the number of errors made in processing trades. Errors at Market Operations take a variety of forms. Money might be posted to the wrong account, or a client might be charged more than the agreed-upon fee. One particularly troublesome error is setting a trade deal on a wrong date, which can create costly problems when the exchange rate or share price changes between the intended date and the erroneous one. Market Operations executives receive monthly incident reports describing those errors that have altered outcomes for the bank by $20,000 or more. Sometimes, the bank loses money due to the error; on other occasions, it makes money (depending on the direction of changes in exchange rates or share prices). Reading through these incident reports, the executives noted that a significant proportion of the errors seemed to be preventable. In addition, some felt that most errors were due to "people" factors such as lack of risk awareness, noncompliance, and poor judgment. Executives who had worked previously in other banks felt that this was a symptom of SBC's "weak risk culture"; that is, they believe that compared to other banks, employees at SBC's Market Operations are more inclined to take irresponsible risks.

1. **How clearly defined is the problem (what, who, when, where, why)?**
- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

2. **Is it clear what the organizational consequences of the problem are?**
- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

3. **Is it clear how serious and urgent the problem is?**
- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

4. **Is it clear what the major cause(s) of the problem could be?**
- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

5. **Is it clear what the logic model is?**
- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

Sub-conclusion 1

Based on the answers to these five questions, you should be able to conclude whether the problem is sufficiently clearly described. When the answers suggest the problem is unclear, there is no point proceeding with the next step. After all, when a problem is unclear – or possibly nonexistent – you cannot solve it, even when you take an evidence-based approach. When the problem is sufficiently clear, you should describe what the problem is, its organizational consequences, its major cause(s), and the PICOC. Use this description as input for step 2, determining whether the problem and (assumed) underlying cause are supported by the evidence.

Step 2: What is the evidence for the problem?

In the previous section, we explicitly used the term assumed problem because we don't yet know what the evidence is to support this problem and/or its underlying cause. Thus, the second step is to ask questions to acquire evidence from multiple sources. How to acquire evidence from practitioners, the scientific literature, the organization, and the most relevant stakeholders is discussed in detail in modules 3, 5, 8, and 10. In this module, we focus on what questions you can ask.

Evidence from Practitioners

The professional judgment of experienced practitioners is an essential component for determining whether an assumed problem is a serious problem, whether the assumed cause is the primary or root cause, and whether alternative causes are plausible. Important questions to ask are

1. Do the practitioners agree with the description of the problem?
2. Do they see plausible alternative causes of the problem?
3. Do they agree that the problem is both serious and urgent?

Evidence from the Organization

Organizational data can be hard, or quantitative, indicators such as staff turnover, error rates, or productivity levels, but they can also include soft elements such as job satisfaction or attitudes toward senior management. This type of evidence includes data from governments, international bodies, and industry bodies. Organizational data are essential to identifying relevant problems and determining possible causes. Important questions to ask are

1. Do the organizational data confirm the assumed problem?
2. Is there a trend? (Do the data suggest the problem will increase if no action is taken?)
3. Do the data confirm the logic model? Is the e a correlation between the assumed cause, the perceived problem, and its organizational consequences?

Evidence from the Scientific Literature

When referring to scientific literature, we mean empirical studies published in peer-reviewed academic journals. In recent decades, a large amount of research has been published on a wide range of managerial issues, such as absenteeism, job satisfaction, improving performance, preventing errors, and motivating employees. Many of these studies also provide insight into the most common causes of these issues. Thus, when tackling these issues in practice, it is important to consult scientific studies. Important questions to ask are

1. Does the scientific literature confirm the assumed major cause of the problem?

2. Does the literature confirm the logic model? (Is the e a correlation between the cause of the problem and its organizational consequences?)
3. Is the evidence generally applicable in the context of the organization (PICOC)?

Evidence from Stakeholders

Stakeholders are individuals or groups who may be affected by an organization's decisions or practices. Internal stakeholders include employees, managers, and board members. However, stakeholders outside the organization, such as suppliers, customers, shareholders, the government, and the public at large, may also be affected. As with evidence from experienced practitioners, evidence from stakeholders is an essential component in determining whether a perceived problem is indeed a serious problem. Stakeholders are also important to understanding whose support may be needed in solving the problem. Important questions to ask are

1. Do they agree with the description of the problem?
2. Do they see plausible alternative causes of the problem?
3. Do they agree that the problem is both serious and urgent?

Learn by doing 2.5

Read the following scenario, which builds on the SBC example used earlier, and determine whether the (assumed) problem is sufficiently supported by evidence from multiple sources.

SBC's executives decide to see whether there is evidence to support their hypothesis that errors are being made because of employees demonstrating lack of risk awareness, noncompliance, and poor judgment – all caused by the firm's weak risk culture.

They hold several focus groups with experienced managers and employees from Market Operations. The participants don't think errors are attributable to a weak risk culture. They report that in the past three years, most of the work within Market Operations that used to be done in London, Frankfurt, and New York is now offshored to low-income branches in Kuala Lumpur, Sao Paulo, and Guangzhou. They point out that employees in these branches are often less experienced and tend to work under poor conditions – factors that increase the risk of errors.

When the executives consult the organizational evidence, they notice that in the past year, only 267 incident reports were filed. The data experts within the bank, however, are certain that the true number of incidents is much larger, probably by a factor 10. Given that within Market Operations, between 30,000 and 40,000 trades are processed every year, this suggest the error rate is rather low. Unfortunately, no data are available that relate directly to the incidents: there is no information regarding the employee(s) involved (e.g., age, tenure/seniority, background/education, experience), nor is there information regarding the branch involved (e.g., geographical location, turnover rate, absenteeism, employee satisfaction, productivity). The data do indicate that the number of errors has significantly increased in the past three years.

Relevant stakeholders say they don't agree the organization suffers from a weak risk culture. They feel that the culture at other banks is not much different and think that the number of errors is not above the average in the sector. In addition, they point out that if the organizational evidence is correct and the number of errors has indeed increased in the past three years, this coincides with the increase in number of offshore locations.

Finally, the evidence from the scientific literature indicates that although the term risk culture is widely used in the financial sector, there is little agreement on what it actually means and whether risk culture can be managed (or changed) to any extent. In addition, it becomes clear that studies on the effect of a company's risk culture on incidents are noticeably absent, and reliable measurement tools are not available. However, studies on errors and incidents in high-risk industries such as aviation, the offshore oil industry, and hospitals clearly demonstrate that the number one predictor of people making mistakes are distractions/task interruptions. In fact, several studies in health care suggest that even fairly simple distractions tend to increase major medical mistakes eightfold.

1. **Does the evidence from practitioners support the assumed problem?**
- To a great extent
- Somewhat
- Very little
- Not at all

2. **Does the evidence from the organization support the assumed problem?**
- To a great extent
- Somewhat
- Very little
- Not at all
- Unclear

3. **Does the evidence from relevant stakeholders support the assumed problem?**
- To a great extent
- Somewhat
- Very little
- Not at all

4. **Does the evidence from the scientific literature support the assumed problem?**
- To a great extent
- Somewhat
- Very little
- Not at all

Sub-conclusion 2

Based on the answers to these questions, you should be able to conclude whether the evidence supports the assumed problem. When the answers suggest that the problem is not supported (or even that it is contradicted) by the evidence, the probability that any solution will effectively address the problem is low, and there is no point proceeding to step 3, the preferred solution.

Step 3: What is the preferred solution?

Aside: The importance of multiple options

Before you ask questions to determine whether a solution is clearly described, it is important to determine whether you have considered more than one solution. The scientific literature suggests that considering multiple solutions tends to lead to better decisions than fixating on yes/no or either/or choices. When we consider only one solution, we tend to ignore evidence contradicting its expected results.

Conversely, considering two or more options leads us to gather more information regarding expected differences in results, which in turn leads to a more informed (evidence-based) decision. When the options are clear, you can ask four follow-up questions:

1. How clearly defined is the preferred solution?

Having a clear description of the preferred solution is a prerequisite to solving the problem, because if you don't have a clear idea of what the solution entails and how it is assumed this course of action will solve the problem, you cannot implement it. A good description entails at least five elements: What? Who? Whe e? When? Why?

2. Is it clear what the logic model is—that is, how the solution would solve the problem?

As explained, a logic model spells out, in a short narrative, the process by which a solution is assumed be the correct action to solve the underlying cause that leads to a problem and its unwanted organizational consequences. For example, it is possible to claim that centralization of administrative functions leads to an efficiency gain of x pe cent, but to have confidence in this solution, you should know ***how*** centralization would lead to this efficiency gain.

In this case, the logic model could be that *centralization of administrative tasks enables* > *standardization of processes* > *which eliminates replication of tasks* > *which reduces costs on labor for duplication of work.* In the Yahoo! example we used earlier, the assumed logic model is that *flexible working arrangements lower performance* > *thus, eliminating all work-from-home arrangements within the company will increase performance* > *which will give the company a competitive edge in the market.*

Note that in this example, it is essential that we first establish that sufficient strong/trustworthy evidence is available to support the assumption that work-from-home arrangements negatively affect performance (evidence from the scientific literature indicates that this is not likely). [3]

3. How clearly defined are the costs and benefits of each solution?

Even the "best" solution may come with considerable costs, so a thorough assessment of the expected costs and benefits is a prerequisite to evidence-based decision-making. There are several analytic tools and templates you can use. Some well-designed tools convert the costs and benefits of each solution into a common unit of measurement (usually money) and then analyze which solution is the most cost efficient. Many analysis tools focused on analyzing costs and benefits, however , do a poor job of identifying indirect and intangible costs (e.g., a decrease in customer satisfaction or drop in employee morale). Thus, when conducting a costs-benefits analysis, it is important that you consult multiple sources of evidence (organizational data, professionals, and stakeholders).

In addition, most analyses fail to attach a degree of uncertainty to the estimated costs and benefits. ithout having an indication of how certain or uncertain the costs or benefits will be, the outcome of your analysis will be misleading. Thus, a good analysis should explicitly factor in the quality of evidence regarding each cost and benefit and attach an estimate of the degree of uncertainty.

4. Is it clear what the best and/or most feasible solution is?

Based on the logic model and costs-benefits analyses of each solution, your organization should have a clear idea of what the best and/or most feasible solution would be.

Learn by doing 2.6

Read the following scenario and determine whether the preferred solution is sufficiently clear.

SBC's executives conclude that the evidence does not support their hypothesis: that errors are being made because employees demonstrating lack of risk awareness, noncompliance, and poor judgment – all caused by the firm's weak risk culture. They conclude it is more likely that most errors in Market Operations are being caused by distractions and interruptions of employees' workflow, which in turn is caused by the poor working conditions at the bank's offshore locations. In particular, they propose to optimize the workplace environment for employees who are at risk of making errors by identifying possible sources of distractions and either eliminating or redesigning them.

In addition, they propose to take actions that have proven to be effective in other industries, such as marking off no-interruption zones; designating distraction-free workplaces; and establishing etiquette rules regarding noise, conversations, phone calls, and other activities that distract employees.

The executives expect that this practical solution significantly reduce distractions and decrease the number of errors. The costs for the solution are expected to be low, and although it is unclear what the direct and indirect costs of the errors and incidents are, the solutions cost are considered a good investment.

1. Were multiple options considered?

- Yes
- No

2. How clearly defined is the preferred solution?

- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

3. Is it clear what the logic model is—how the solution would solve the problem?

- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

4. **Is it clear what the solution's costs and benefits are?**
- Very clear
- Fairly clear
- Somewhat unclear
- Very unclear

Sub-conclusion 3

Based on the answers to these questions, you should be able to conclude whether the preferred solution is sufficiently clearly described. Again, when the answers suggest that the preferred solution or its logic model is unclear, there is no point proceeding with the next step. When the solution is sufficiently clear, you should describe in detail what the preferred solution is, its logic model, and the PICOC. Use this description as input for step 4, determining whether the preferred solution is supported by the evidence.

Step 4: What is the evidence for the solution?

As was the case with the assumed problem, the next step is to ask questions to determine whether the evidence supports the assumed effectiveness of the preferred solution. Again, how to acquire evidence is described in modules 3, 5, 8, and 10. This module focuses on what questions you can ask.

Evidence from Practitioners

The professional judgment of experienced practitioners inside and outside the organization is an essential component in determining how likely a proposed solution is to work in a particular context. In addition, experienced professionals are often in a good position to rate the solution in terms of implementation costs and other feasibility and risk issues. Finally, experienced professionals may think of alternative solutions that you haven't considered. Important questions to ask are

1. Do they agree on which solution is the best and/or most feasible?
2. Do they see downsides to or unintended negative consequences of the preferred solution?
3. Do they see alternative solutions to the problem that may work better?

Evidence from the Organization

Ideally, you would have organizational data available that could help determine which solution has the highest likelihood of solving the problem. If this is the case, an important question to ask is

1. Can organizational data be used to monitor the future effectiveness of the preferred solution?

Evidence from the Scientific Literature

As explained earlier, a lot of research has been published on a wide range of managerial issues, such as improving performance, preventing errors, and motivating employees. Many of these studies also provide insight into which variables or management interventions may have a positive impact. Thus, when it comes to tackling these issues in practice, it is important to consult these studies. Important questions to ask are

1. What does the scientific literature suggest regarding the effectiveness of the preferred solution?
2. Does the literature suggest other solutions to the problem that may work better?
3. Is the evidence generalizable to the organizational context (PICOC)?

Evidence from Stakeholders

Even the best solution can fail upon implementation if the stakeholders see serious downsides or if they feel an alternative solution may work better. Gathering evidence from stakeholders is therefore an essential component in determining how likely a proposed solution is to work in a particular context. In addition, stakeholders are often in a good position to judge the preferred solution in terms of implementation costs and other feasibility and risk issues. Finally, stakeholders may see alternative solutions that you haven't considered. Important questions to ask are

1. Do they agree on which solution is the best and/or most feasible?
2. Do they see downsides to or unintended negative consequences of the preferred solution?
3. Do they see alternative solutions for the problem that may work better?
4. Are they supportive of the preferred solutions?

Learn by doing 2.7

Read the following scenario and determine whether the preferred solution is sufficiently supported by evidence from multiple sources.

SBC's executives decide to see whether there is evidence to support the preferred solution: optimizing the workplace environment for employees who are at risk of making errors by identifying possible sources of distractions and either eliminating or redesigning them. Again, they hold focus groups with experienced managers and employees from Market Operations. All agree that interventions to reduce or eliminate distractions and task interruptions will most likely reduce the number of errors. In addition, they feel that solutions that have proven to be effective in other industries will also be applicable to the organizational context of SBC Market Operations. Moreover, several managers indicate that they have experience with some of the suggested interventions as implemented at other companies, and they confirm that those interventions significantly decreased the error rate.

The bank's data experts agree to systematically collect data regarding the number of errors as well as to capture relevant information regarding both the employees and the locations involved. These data will be analyzed and summarized on a monthly basis and made available to the managers and executives.

When the scientific evidence is consulted, it becomes clear that several high-quality studies indicate that the most common distractions come from coworkers. In fact, face-to-face interruptions account for one-third more intrusions than email or telephone calls, which employees feel freer to ignore. As a result, employees in open-plan offices – which are widely used at the offshore branches in Kuala Lumpur, Sao Paulo, and Guangzhou – are interrupted an estimated 20 to 30 percent more often than those in private offices – which are more common at the bank's branches in London, Frankfurt, and New York. In addition, it was found that even simple interventions in the office design or establishing etiquette rules significantly reduces the number of distractions and task interruptions.

Finally, relevant stakeholders agree that the preferred solution will most likely help reduce the number of errors within Market Operations. In particular, employees who work at the offshore locations are very pleased with the proposed interventions and expect that their working conditions will significantly improve. The managers responsible for the offshore locations are certain that the preferred solution will lower the operational costs in the long term and are therefore happy to actively support its implementation.

1. **Does the evidence from practitioners support the preferred solution?**
- To a great extent
- Somewhat
- Very little
- Not at all

2. **Can organizational data be used to monitor the future effectiveness of the preferred solution?**
- Yes
- No

3. **Does the evidence from the scientific literature support the preferred solution?**
- To a great extent
- Somewhat
- Very little
- Not at all

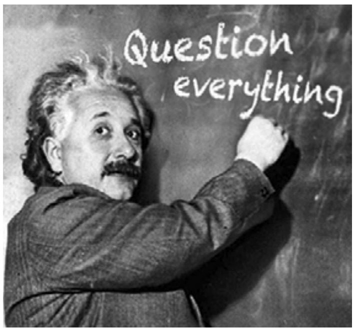
4. **Does the evidence from relevant stakeholders support the preferred solution?**
- To a great extent
- Somewhat
- Very little
- Not at all

Final Conclusion

Based on the answers to these questions, you should be able to conclude whether the evidence supports the decision regarding the best and/or most feasible solution. Again, when the answers suggest that the preferred solution is not supported (or even that it is contradicted) by the evidence, the likelihood that the solution will effectively address the problem is low. In that case, you are left with only one option: to go back to the drawing board.

In addition, if you conclude that the available evidence is too limited, you should acquire additional evidence. Only when sufficient (strong/trustworthy) evidence supports the potential effectiveness of the preferred solution would you consider the decision to implement to be an ***evidence-based*** decision.

Developing your capacity to ask questions



As stated at the beginning of this module, asking questions kicks off the deliberate search for evidence. In fact, stopping the search process too soon often leads to solving the wrong problem or settling on a solution before really understanding its possible (side) effects. In a rush to get things done, the question-asking phase – the first step of evidence-based management – may be suppressed and uncertainty glossed over. The resultant lack of evidence and insight leads to solving the wrong problem or pursuing a questionable solution with limited results. Unfortunately, there is no shortcut to a thoughtful start to the evidence-based process, regardless of urgency or the resources poured into a problem. The more time spend assessing the problem, the less time required to solve it. This means asking questions – lots of them – to check assumptions, particularly where someone (including ourselves) asserts a belief as a certainty.

This habit-forming approach can inform your conversations and deliberations. You will begin to ask yourself and others, “What’s the evidence for that?”, as impressions, beliefs, and attitudes appear in your conversations about the organization, its practices, and the decisions being made. This approach has turned many of our students and course members into the “evidence squad,” and they learn to use it over time in a manner that promotes asking critical questions about evidence without necessarily criticizing. Concern for the evidence behind decisions translates into active questioning and healthy skepticism. Evidence-focused questioning of claims, statements, or assertions changes both the conversations and deliberations of emergent evidence-based managers.

A must here is for practitioners to learn ways to raise these questions in socially effective ways (read: *civil and persuasive*). To be effective, evidence-based managers need to avoid being dismissed as mere naysayers. Raising questions can be anxiety-provoking for would-be evidence-based managers, who fear making waves. This questioning extends to assertions made by professors, consultants, and other experts. So, yes, we expect you to question your lecturers by critically considering their arguments and reviewing their sources. Once practiced at it, evidence-based managers become comfortable at asking, “Is this your personal opinion based on your own professional experience, or is there any evidence in support of it?” You may be surprised to learn how much uncertainty really exists regarding the practices your organization uses.

An evidence-based approach begins with asking questions – lots of them – to identify assumptions underlying a claim, assertion, or hypothesis regarding a problem and/or solution and to check whether sufficient evidence supports these assumptions. Of course, managers, policymakers, and business leaders make assumptions all the time, so your first task as an evidence-based practitioner is to identify the most critical assumptions: assumptions that, if ultimately false, have severe negative consequences for the decision at hand.

The next key point is that the problem you are trying to solve should be clear. An organization may learn about a new management technique, assume that the company somehow will benefit f om it, and thus decide to implement this new solution. However, implementing a solution when there is no evident problem can be a serious waste of time and resources. For this reason, an evidence-based approach always starts with the question, “What is the problem you are trying to solve, and what is the evidence for this problem?”

We brie y discussed a conceptual tool to help you find evidence that takes into account your professional context: PICOC (Population, Intervention, Comparison, Outcome and Context).

Questions regarding the (assumed) problem

When you have a satisfactory answer regarding the most important problem, you should ask these five follow-up questions:

1. How clearly defined is the problem (who, when, where, what, why)?
2. Is it clear what the organizational consequences of the problem are?
3. Is it clear how serious and urgent the problem is?
4. Is it clear what the major cause(s) of the problem could be?
5. Is it clear what the logic model is?

Based on the answers to these five questions, you should be able to conclude whether the problem is sufficiently clear. You can use this problem description as input for the next step: determining whether the problem and (assumed) underlying cause is supported by the evidence. Important questions to ask are

- Do experienced practitioners and relevant stakeholders agree with the description of the problem? Do they see plausible alternative causes of the problem? Do they agree that the problem is both serious and urgent?
- Do the organizational data confirm the assumed problem? Is there a trend? Do the data suggest the problem will increase if no action is taken?
- Does the scientific literature confirm the assumed major cause of the problem? Is there a correlation between the cause of the problem and its organizational consequences?

Questions regarding the (preferred) solution

It is important to determine whether more than one solution is considered. Considering two or more options leads to gathering more information about expected differences in results and yields a more informed (evidence-based) decision. When it is clear what the options are and which of those options is the preferred solution, you should ask these four follow-up questions:

1. How clearly defined is the preferred solution (who, when, where, what, why)?
2. Is it clear what the logic model is – that is, how the solution would solve the problem?
3. How clearly defined a e the costs and benefits of each solution
4. Is it clear what the best and/or most feasible solution is?

Based on the answers to these four questions, you should be able to conclude whether the preferred solution is clearly described. You can then use the description as input for the next step: determining whether the preferred solution is supported by the evidence. Again, important questions to ask are

- Do experienced practitioners and relevant stakeholders agree on which solution is the best and/or most feasible? Do they see downsides or unintended negative consequences? Do they see alternative solutions to the problem that may work better?
- Can organizational data be used to monitor the future effectiveness of the preferred solution?
- What does the scientific literature suggest regarding the effectiveness of the preferred solution? Does the evidence suggest other solutions to the problem that may work better? Is the evidence generalizable to the organizational context (PICOC)?
- Are the most relevant stakeholders supportive of the preferred solution?

Based on the answers to these questions, you should be able to conclude whether the evidence supports the decision (i.e., which solution is best and/or most feasible). Remember that when the solution is not supported (or even contradicted) by the evidence, its likelihood of success is low. You are then left with only one option: go back to the drawing board.

Finally, Never Stop Asking . . .

Last, we discussed the importance of developing your capacity to ask critical questions. Asking questions kicks off the deliberate search for evidence. Stopping the search process too soon may lead to solving the wrong problem or implementing an ineffective solution. You need to learn ways to raise critical questions in a ***socially effective*** way. With practice, you'll become skilled at asking in a civil but persuasive way: "Is this your personal opinion based on your own professional experience, or is there any other evidence in support of it?" Sometimes, it's challenging to speak up when the available evidence does not support the solution favored by powerful executives, but remember this: asking critical questions is the lifeblood of evidence-based management.

Podcast

In this podcast host Karen Plum discusses with Denise Rousseau, Professor of Organizational Behavior and Public Policy at Carnegie Mellon University, Eric Barends, Managing Director of the Center for Evidence Based Management (CEBMa), and Lisa Grififths, CEO of OzChild (an Australian child welfare organization) the importance of asking questions. Critical questions, to understand what the problem is that we are trying to solve. This is because so often in organisations we leap to solutions without being clear what the problem is.




This approach can be driven by a desire to do “stuff”, to get things done and to feel that we are making progress. But how many times is a solution delivered and it doesn’t seem to do the trick? Maybe that’s because not enough time was taken to understand the nature of the problem before jumping to solutions.

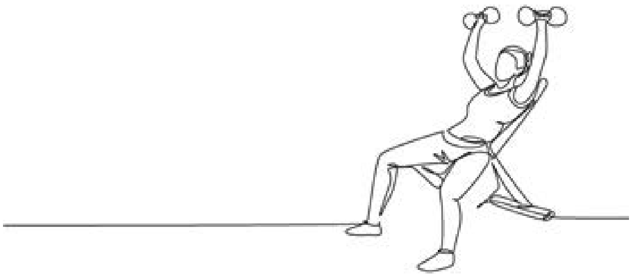
So we need to ask questions – lots of them. Asking them in the right way, of the right people, at the right time is vital.

By being constructive in our questioning, we can make a contribution to the decision-making process even if we are not experienced, not powerful or not senior in the organisation. Sometimes the inexperienced have a better perspective because they don’t make so many assumptions and will be more easily forgiven for asking the “daft questions”.

In other words “question everything”!

https://evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 2.1: Surfacing & checking assumptions

Take a policy paper, project proposal, strategy document or change plan your organization uses. If you are not able to obtain one, you can download one ***here >>***

1. *Read (skim) the paper.*
2. *Determine two important 'core' assumptions - assumptions that, if incorrect, will have a major impact on the desired outcome.*
3. *Ask the author(s) of the document - in a respectful and non-threatening way - what the evidence is for these two assumptions.*
4. *Do you feel that the evidence provided is sufficient in both quantity and quality to support the assumptions? Explain why.*
5. *If you feel insufficient evidence is available, what additional evidence should be obtained and from which sources?*

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Exercise 2.2: Lagging productivity – the case of Psy-Care

Imagine you are an (evidence-based) consultant for a large mental health care organization (Psy-Care). The organization's productivity appears to be below set targets agreed upon with the health insurance companies funding the organization. If at the end of the year the targets are not met, the insurance companies will cut the organization's funding and as a result jobs will be lost. For this reason, the board of directors has prepared a letter they intend to send to all of Psy-Care's health care professionals. A copy of the letter can be downloaded ***here >>***.

From this letter, it appears that the board of directors assumes that the organization's productivity targets are not being met because therapists don't see enough patients. Their logic model looks like this:

Therapist don't see enough patients > as a result fewer billable hours are logged into the organization's registration system > therefore the agreed-upon number of billable hours will not be met > consequently the insurance companies will cut the organization's funding > to reduce costs, the Psy-Care needs to cut jobs.

1. *Describe in detail what evidence and from which sources you would obtain in order to verify the board of directors' claim that Psy-Care's productivity is lagging because therapists don't see enough patients.*
2. *Describe four alternative explanations (i.e., underlying causes) for the low number of billable hours in the system.*
3. *Select two alternative causes and describe for each cause the evidence you would want in order to verify it:*
   - *What would you ask the managers and the stakeholders?*
   - *What scientific research evidence would you look for?*
   - *What organizational data would you obtain?*
   - *What other evidence sources would you consult?*

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor

Suggestions for further reading

- *Before you make that big decision*, Kahneman et al, Harvard Business Review, June 2011.
- *Cross Section, Interview with Neil deGrasse Tyson*, Science Weekly podcast, The Guardian, 2016.
- *The surprising power of questions*, Brooks & John, Harvard Business Review, May-June 2018.

References

1. The Guardian. (December, 2016). *Cross Section: Neil deGrasse Tyson* (interview with Nicola Davis). Science Weekly Podcast.
2. Saber Tehrani, A. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). “25-Year summary of US malpractice claims for diagnostic errors 1986–2010: An analysis from the National Practitioner Data Bank.” *BMJ Quality & Safety*. Volume 22. Number 8. 672-680.
3. De Menezes, L. M., & Kelliher, C. (2011). “*Flexible working and performance: A systematic review of the evidence for a business case.*” International Journal of Management Reviews. Volume 13. Number 4.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 2.1

1: D

2: A > You could rightly argue that the claim that Apple never spends money on marketing research is false. However, that is not the underlying assumption.

3: B > You could rightly argue that the claim that 3M allows their employees to use up to 15 percent of their time to pursue new, innovative ideas is false, but that is not the underlying assumption.

Did I get this? 2.1

C >The underlying assumption is that a low absenteeism rate is the direct result of the organization's HR department.

Did I get this? 2.2

A and B

Learn by doing 2.2

Our answer: We would argue that obvious possible causes are time-consuming bureaucratic procedures, poor logistics, or inefficient operational p ocesses. But if you think out of the box, a wide range of underlying problems are possible. For example:

1. The nurses at the hospital's emergency ward are undereducated and often fail to recognize the symptoms of a stroke.
2. The nurses are undereducated because the hospital pays the lowest salary in the region; consequently, highly skilled nurses go elsewhere.
3. The IT system at the hospital's emergency ward is obsolete; consequently, there are many malfunctions, and stroke patients just disappear from the system.
4. The hospital's procurement department does not function efficiently; consequently, the e is often a shortage of thrombolytic drugs, and a courier must be sent to a neighboring hospital.
5. The hospital's procurement department does not function efficiently because it is understa fed; in this part of the United Kingdom, it is hard to get skilled procurement professionals.
6. The hospital's architecture and layout is too complex and very confusing; consequently, stroke patients who are referred to another department for thrombolytic therapy often lose their way and arrive much later than planned.

Learn by doing 2.3

P = Canadian college students

I = Financial incentives

C = Performance goals

O = Academic achievement

C = A Canadian community college

Did I get this? 2.3

P = Millennials

C = Doing nothing – business as usual

O = A better work–life balance

Learn by doing 2.4

1. Somewhat/very unclear: It is clear that errors are being made in SBC's Market Operations, but we would argue that it is unclear how many errors are made and how much money is involved. In addition, the executives' assumption that this is due to people factors and a weak risk culture is rather vague.
2. Somewhat/very unclear: We don't know what the organizational consequences of the errors in Market Operations are (or might be), so this is a question we must ask and address.
3. Very unclear: If we don't know what the organizational consequences are (or might be), we also don't know how serious and urgent the problem is, so we must ask about and address this issue: What will happen if nothing is done?
4. Somewhat/very unclear: The executives' assumptions that errors are made due to people factors and that its major cause is the bank's weak risk culture lacks clarity, so we must ask about and address this issue.
5. Somewhat/very unclear: Important elements of the logic model are missing. The executives assume that a weak risk culture leads to employees demonstrating irresponsible risk taking, poor compliance, and bad judgment, which in turn lead to errors. It is unclear, however, whether (and how) the errors lead to organizational consequences and whether (and how) these consequences affect the company's business outcomes.

Learn by doing 2.5

1. Very little / not at all: The outcome of the focus groups indicates they don't think errors are attributable to a weak risk culture. In fact, they see an alternative cause for the problem: the limited experience and poor working conditions of employees at the offshore locations.
2. Unclear: There is little relevant organizational data available, so the evidence neither supports nor contradicts the assumed problem.
3. Very little / not at all: The stakeholders don't agree the organization suffers from a weak risk culture. They feel that the culture at other banks is not much different. In addition, they see an alternative cause for the problem: the increased number of offshore locations.
4. Very little / not at all: The scientific literatu e indicates that there is little agreement on what the term risk culture means and whether it can be managed (or changed). As a result, studies on the effect of a company's *risk culture* on incidents are noticeably absent. In addition, the evidence suggests an alternative cause for the problem: distractions.

Learn by doing 2.6

1. No > Only one solution was considered: optimizing the workplace environment for employees who are at risk of making errors.
2. Very clear > The solution to optimize the workplace environment for employees who are at risk of making errors is illustrated with clear practical examples, such as marking off no-interruption zones, designating distraction-free workplaces, and establishing etiquette rules regarding activities that distract employees.
3. Very clear > The assumed logic model is that *poor working conditions lead to distractions and task interruptions> which lead to preventable errors > thus, optimizing the workplace environment will lead to fewer distractions > and consequently to fewer errors > which in turn will reduce the direct and indirect costs associated with errors.*
4. Fairly clear > It is unclear what the direct and indirect costs of the errors and incidents are, but the costs for the preferred solution are expected to be low, so it is considered a good investment.

Learn by doing 2.7

1. To a great extent > The outcome of the focus groups suggests that practitioners agree that interventions to reduce or eliminate distractions and task interruptions will most likely reduce the number of errors. Moreover, several managers indicate that they have experience with some of the suggested interventions as implemented at other companies, and they confirm that thos interventions significantly dec eased the error rate.
2. Yes > The bank's data experts have agreed to systematically collect data regarding the number of errors as well as to capture relevant information regarding both the employees and the locations involved. These data can be used to monitor the solution's effectiveness.
3. To a great extent > The research literature indicates that even simple interventions in the office desig or establishing etiquette rules significantly educes the number distractions and task interruptions.
4. To a great extent > Stakeholders agree that the preferred solution will most likely help reduce the number of errors. In addition, managers responsible for the offshore locations are happy to actively support the implementation.

CEBMa center for Evidence-Based Management

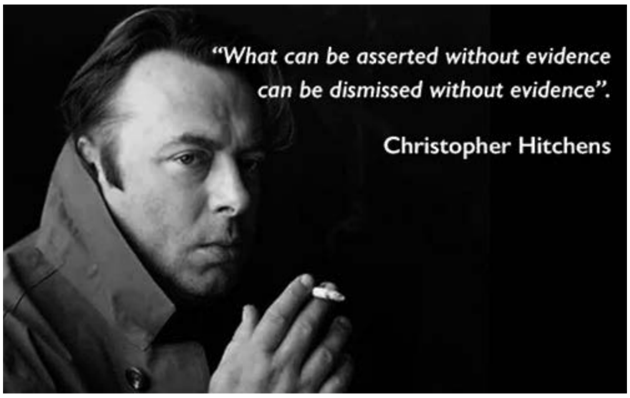
Open Learning Initiative

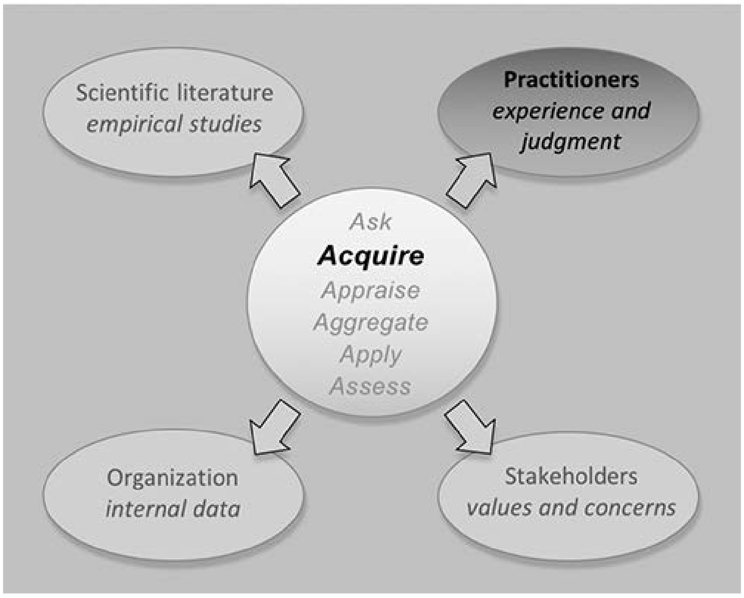
MODULE 3 | ACQUIRE Evidence from Practitioners

Learning objective:

- Acquire evidence from practitioners in a valid and reliable way.

Introduction





In organizations, evidence from practitioners is an important source of information. It may well be the most-used source of information in decision-making processes. In many organizations, leadership teams ask employees for their input, managers for their opinions, and consultants for their experience, and managers often base their decisions on this type of evidence. The quality of decisions significantly improves when you consider all sources of evidence, including professional expertise, as it can connect external evidence such as scientific research findings to the specific organizational context.

There are many ways in which you can acquire evidence from practitioners. Numerous books and websites are available that can inform you about how to gather evidence in a valid and reliable way, covering important aspects such as sampling procedures, research designs, and questionnaire development. This module therefore does not aim to give a comprehensive and detailed overview of all methods in which you can acquire evidence from practitioners. Rather, it provides a quick summary of key aspects that you should consider.

What to ask

Acquiring evidence from practitioners is not a fishing expedition – it starts with an assumed problem, a preferred solution, or a deemed opportunity. Before contacting practitioners to ask about their take on the matter, it is therefore important that you first clearly describe the (assumed) problem that needs to be solved or the opportunity that needs to be addressed. In general, a good definition of the problem entails at least three elements:

1. The problem itself, stated clearly and concisely (What? Who? Where? When?)
2. Its (potential) organizational consequences
3. Its assumed major cause(s)

The professional judgment of experienced practitioners inside and outside the organization is an essential component in determining whether the assumed problem is indeed a serious problem and in identifying possible causes. Thus, important questions to ask are

1. Do you agree with the description of the problem?
2. Do you see plausible alternative causes of the problem?
3. Do you agree that the problem is serious and urgent?

In addition, when you need to make a decision that involves whether to implement a proposed solution, having a clear description of that solution is a prerequisite before you consult practitioners. A good description of a solution entails at least three elements:

1. The solution itself, stated clearly and concisely (What? Who? Where? When?)
2. Its (potential) effect on the problem and underlying causes
3. Its costs and benefit

As explained, evidence from practitioners is also an essential component in determining how likely a proposed solution is to work in a particular organizational context. In addition, experienced professionals are often in a good position to rate the preferred solution in terms of implementation costs and other feasibility and risk issues. Finally, experienced professionals may think of alternative solutions that you haven't considered. Thus, important questions to ask are

1. Do you agree on which solution is the best and/or most feasible?
2. Do you see downsides to or unintended negative consequences of the preferred solution?
3. Do you see alternative solutions to the problem that may work better?

Whom to ask

The first step in gathering evidence f om practitioners is determining the target audience. Which practitioners in the organization are, given the question or issue at hand, most likely to provide a valid and reliable judgment? Obviously, we want practitioners whose professional judgment is based on a high level of expertise. ***Expertise*** refers to skill and knowledge acquired through training and education coupled with prolonged practice in a specific domain and combined with frequent and direct feedback. So, for example, if the question concerns a proposed solution for high levels of staff absenteeism among lawyers in an international accounting firm, then a practitioner who has brief experience only with absenteeism among deli workers at a small supermarket chain should obviously not be the first one whose professional judgment is sought. The quality or trustworthiness of practitioner expertise and judgment depends on the relevance of the practitioner's training, education, and experience, which we discuss in Module 4.

Sample size: How many practitioners should I ask?

In most cases, it is impossible to ask all practitioners in the organization to give their judgment, so we need a sample – a selection of practitioners chosen in such a way that they represent the total population. But how many practitioners should your sample consist of? Should you ask 1 percent, 5 percent, 10 or even 50 percent of the practitioners in the organization? This depends largely on how accurate you want your evidence to be. Most researchers use a sample size calculator to decide on the sample size. The required sample size, however, also depends on practical factors such as time, budget, and availability. In addition, qualitative methods (e.g., focus groups involve a substantially smaller sample size than quantitative methods (e.g., surveys because, for these methods, representativeness is often more important than accuracy.

Selection bias: Which practitioners should I ask?

Another major concern is selection bias. Selection bias, also called sampling bias, occurs when your selection of practitioners leads to an outcome that is different from what you would have gotten if you had enrolled the entire target audience.

Example

n 1936, Democrat Franklin Roosevelt (left and Republican Alf Landon (right) were running for president. Before the election, the magazine Literary Digest sent a survey to 10 million Americans to determine how they would vote. More than 2 million people responded to the poll; 60 percent supported Landon. The magazine published the findings and predicted that Landon would win the election. However, Roosevelt defeated Landon in one of the largest landslide presidential elections ever.

What happened?

The sample systematically underrepresented Democrats, and the result was a whopping error of 19 percent, the largest ever in a major public opinion poll. [1]

Learn by doing 3.1

Read the following scenario:

A company decides to survey its employees to understand their satisfaction with the new flexible working hours. On Monday at 9 a.m., a survey questionnaire is administered to the first 200 employees who start their working day.

A. Which of the following might negatively affect the outcome of the survey? Check all that apply.

1. A sample of 200 is too small.
2. Monday at 9 a.m. is not a good moment to administer a survey.
3. The sample is not random.

B. What would be the best time to distribute a questionnaire during a three-day conference?

1. At the start of the conference
2. On the second day of the conference
3. At the end of the conference

How to ask

Walk around and ask

The quickest and easiest way to gather evidence from practitioners is by walking around and asking. Of course, this method is prone to bias, but sometimes wandering around in an unstructured manner through the workplace, and asking people – randomly – their judgment about an assumed problem or preferred solution is a good way to start.

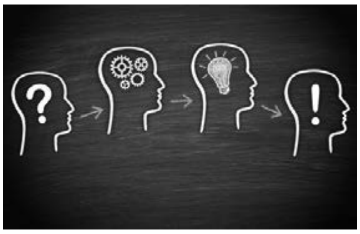
Conduct a Survey

A survey is a quick and efficient way to ask a large group of people a question (or a series of questions) to gather evidence about their opinion, judgment, or attitude toward an assumed problem or preferred solution. Participants in a survey are usually selected in a way that the results are generalizable to a larger population. Most surveys are quantitative in nature, meaning that they place more emphasis on numerical than on narrative data, and are therefore intentionally narrow and specific. For this reason, surveys typically present closed-ended questions that provide a list of predetermined responses from which participants can choose their answers.

One of the most common formats used in survey questions is the agree/disagree format. In this type of question, respondents are asked whether they agree or disagree with a particular statement. A better practice, however, is to use Likert scales. An example of a closed-ended survey question using a Likert scale is, “Please rate how strongly you agree or disagree with the following statement: *I receive too many emails*. Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?” An example of an open-ended question is, “What do you think is the most important cause(s) of the high level of absenteeism among the lawyers in our firm?”

NOTE: When surveying practitioners, it is important to inform them in advance about why you need their input. You also need to set clear expectations regarding anonymity, confidentiality, and how information will be used, as this disclosure increases the likelihood of honesty. You can administer surveys in several ways, such as through mail or email, by telephone, or face-to-face. In addition, there are several (free) online survey tools available, of which Survey Monkey is probably the most well-known.

Use the Delphi method



The Delphi method is a qualitative, interactive method entailing a group of experts or professionals who anonymously reply to a questionnaire or a set of statements and subsequently receive feedback in the form of a group response, after which the process is repeated. The method is based on the principle that judgments from a group of individuals are more accurate than those from a single person or an unstructured group, provided that the judgments are made independently before being combined. [2] [3] [4] [5]

The method is meticulously structured and typically involves the following steps:

1. The facilitator develops a questionnaire.
2. The participants independently and anonymously answer the questionnaire.
3. The facilitator summarizes the responses and develops a feedback report.
4. The participants evaluate the feedback report and revise earlier answers that may have changed in light of the replies of other participants.
5. The process is stopped after a predefined criterion (e.g., number of rounds, consensus, stability of answers).
6. The facilitator develops a final summary

The goal of this method is to reduce the range of responses and converge toward expert/professional consensus. The Delphi method has been widely adopted and is used by a broad range of organizations across industries, including public policymaking.

Use a group decision room

A group decision room, also known as an *acceleration chamber* or a *brainbox*, is the electronic version of a focus group or the Delphi method. Participants gather in a meeting room with electronic tools (computer, laptop, or tablet) or connect remotely through the Internet. Under the guidance of an experienced moderator, participants can anonymously respond to questions and/or statements, vote on issues, or react to topics discussed. The responses are visible (in real time) for all participants, so the final outcome can be discussed immediately

Use a mobile voting/audience response system

Mobile voting, or audio response, systems are developed to create life interaction between a presenter/moderator and the audience. In educational settings, such systems are often called *student response systems*. Meeting participants can anonymously vote, answer questions (open ended as well as multiple choice), or give their judgment over the Internet with any smartphone, tablet, or computer. Results are instantly tabulated via the Internet and presented on screen in real time for analysis discussion. Results can also be imported into a database for further evaluation or comparison with other sessions. There are several mobile voting systems available, such as Socrative and VoxVote.

Did I get this 3.1

To which of the following biases would walking around and asking be susceptible? Select Yes or No for each option.

a) Authority bias

b) Confirmation bias

c) Group conformity

d) Selection bias

Did I get this 3.2

Which biases are the Delphi method, a mobile voting system, and a face-to-face meeting prone to? Select Yes or No for each option.

1. Delphi method

a) Authority bias

b) Confirmation bias

c) Group conformity

d) Selection bias

2. A group face-to-face meeting

a) Authority bias

b) Confirmation bias

c) Group conformity

d) Selection bias

3. Mobile voting system

a) Authority bias

b) Confirmation bias

c) Group conformity

d) Selection bias

Developing questions

Perhaps the most important part of gathering evidence from practitioners is the development of questions that accurately measure their opinion, experience, or judgment. Regardless of the method used, the outcome is useless if the evidence gathered results from ambiguous or leading questions. Formulating questions is a process that requires attention to many details. The choice of words in a question is critical to express its meaning and intent – even small wording differences can substantially affect the answers people give. Here are a number of tips to reduce measurement and comprehension errors resulting from the question's wording [6] [7]

8 Tips for writing valid and effective surveys

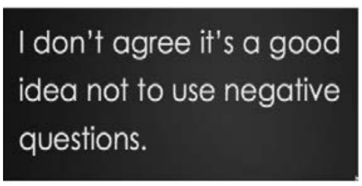
1. Keep it simple

Although you may feel tempted to build beautiful phrases, questions that are short and that use simple and concrete words are more easily understood. Academic language (e.g., cerebrovascular accident instead of stroke) and unfamiliar abbreviations or jargon (e.g., ROI instead of return on investment) can result in respondent confusion and should be avoided. The same is true for complex sentences. For example, questions such as the following are far too complex to answer: "Do you agree or disagree that, controlling for inflation, your income has g own in the last year, where income means your gross household income calculated as the total financial receipts of all adults living in your household?"

2. Avoid double-barreled questions

Make sure you ask only one question at a time. Questions that ask respondents to evaluate more than one concept are often called *double-barreled questions*. An example is "How organized and interesting was the meeting?" If someone answers "moderately" to this question, what does that mean? [2] Moderately organized and moderately interesting? Extremely interesting but only slightly organized? An obvious remedy to this confusion is to write two questions instead of one: (a) "How organized was the meeting?" and (b) "How interesting was the meeting?"

3. Avoid negative, especially double negative, questions

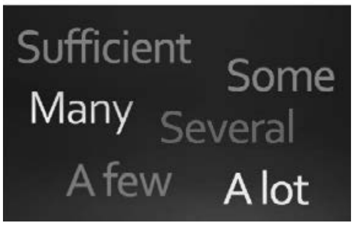


For example, "Do you agree that it is not a good idea not to implement the new IT system?" will probably baffle respondents. The double negative "not a . . . not to" causes this confusion. The same often holds for questions worded in the negative. For instance, how would you interpret an answer to the question "Should the HR director not be directly responsible to the CEO?"

4. Avoid vague or ambiguous terms

Words such as *often*, *regularly*, *sometimes*, *normal*, *substantial*, *might*, *could*, and *probably* should be avoided. For example, “Do you agree that errors are often being made in the accounting department?” is vague, whereas “*How many times per week* do you think errors are being made in the accounting department?” is specific. Don't leave anything to interpretation by respondents. Also avoid abstract terms such as *moral*, *decent*, and *appropriate*, because they have different meanings for different people. Instead of asking, “Do you agree or disagree that moral values are an important issue in our organization?”, ask about specific issues that fall under the broad category of moral values. For example, “Many issues that involve moral values are prominent in our organization today. Below we have a list of these issues. For each issue, please tell us how important you think it is: . . .”

5. Define terms very specifically



In the question “What was your income last year?”, the term *income* is vague. Does it mean personal income or household income? Does it mean income before or after tax? The term *income* therefore needs more specification. For example, “What was your total household income before tax in the past year?”

6. Avoid loaded, leading, or emotional language

The question “Do you agree that the organization should immediately stop the failing implementation of the poorly designed IT system?” contains biasing language: “*immediately* stop,” “*failing* implementation,” and “*poorly* designed.” All of these terms can bias respondents toward a certain point of view. These terms should therefore be omitted or replaced with more balanced language.

7. Prevent social desirability bias

Another challenge in developing questions is social *desirability bias*. Respondents have a natural tendency to want to be accepted and liked, which may lead them to provide “socially desirable” answers, especially to questions that deal with sensitive subjects such as leadership style, accountability, and ethical issues. Research indicates that social desirability bias is more likely to occur when an interviewer is present (e.g., during a face-to-face meeting or telephone survey) than when respondents answer the questions in private (e.g., paper and Web surveys). The best option is, of course, to enable respondents to answer questions anonymously.

8. Pilot test your questionnaire

Finally, it is important to test your questions before using them to acquire evidence. Pilot testing your questionnaire using a small sample of people from the target population helps you identify ambiguities or questions that are unclear. In addition, you get feedback and an estimate of how much time it will take people to respond to your questions.

Did I get this 3.3

1. **Are the following four agree/disagree questions adequately formulated? Select Yes or No.**

a) Roles and responsibilities for managing risk in my area are clear and consistent.

b) The new procedure that was implemented last month is good at preventing errors from happening.

c) In the team meetings I attend, risk management is a regular agenda item.

d) In our department, there is an open environment that is receptive to challenge and improvement.

2. **The following two agree/disagree questions are not adequately formulated. Select the most important shortcoming in each.**

1. We can reduce the number of incidents by learning from them.

a) The question contains vague and ambiguous term.

b) This is a negative question.

c) This is a double-barreled question.

d) The question may lead to social desirability bias.

2. When I report an incident, I get to know the outcome. Feedback is effective.

a) The question contains vague and ambiguous terms.

b) The question may lead to social desirability bias.

c) This is a negative question.

d) This is a double-barreled question.

3. **Which of the following may be a reason this question is not adequately formulated? Check all that apply.**

It can be difficult to find out about changes to regulations and how they affect me.

a) This is a double-barreled question.

b) The question contains leading and emotional language.

c) The question contains vague and ambiguous terms.

d) The question may lead to social desirability bias.

4. **Read the following four questions and determine whether they are adequately formulated. If not, explain how they could be improved.**

a) Sometimes I think about not raising a problem because I worry about how it will be perceived.

b) My line manager is very conscious of risk and shows strong leadership in ensuring we manage risk in the right way. He or she is an excellent role model for me.

c) When this department gets really busy, colleagues from other departments help out

d) In my experience, people in Operations demonstrate a commitment to completion, accuracy, and timeliness. There is careful attention to detail.

To sum up...

We started by discussing how evidence from practitioners can be acquired in a valid and reliable way. The first step is to determine whom to ask. Given the question or issue at hand, which practitioners in the organization are most likely to provide a valid and reliable judgment? And how can selection bias be avoided?

The second step is determining how to ask. The quickest and easiest way to gather evidence from practitioners is by walking around and asking. This method, however, is especially prone to bias. Other approaches include surveys, the Delphi method, a group decision room, or a mobile voting system.

In addition, we discussed the importance of posing questions that accurately measure the practitioner's opinion, experience, or judgment. After all, the outcome is useless if the evidence gathered results from ambiguous or leading questions. We therefore ended this module by providing several tips, among them the following, to reduce measurement and comprehension errors caused by the wording of the question:

- Keep it simple.
- Avoid double-barreled questions.
- Avoid negative questions or emotional language.
- Avoid vague words and define terms specifically
- Prevent social desirability bias.
- Always pilot test your questionnaire.

Podcast

In this podcast host Karen Plum discusses with Denise Rousseau, Professor at Carnegie Mellon University, Eric Barends, Managing Director of CEBMa, Christina Rader, Professor at Colorado College, and Lisa Griffiths, CEO of OzChild how to acquire evidence from practitioners – people who have experience with the problem we are tackling, or the solutions we are considering.

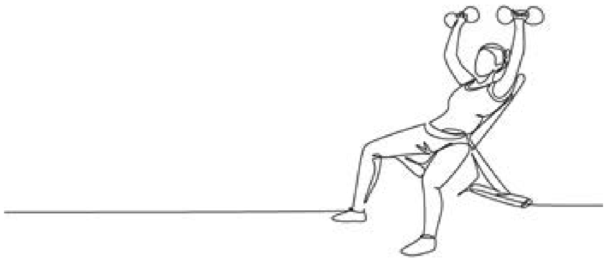
In this podcast, we consider the choice of practitioners (people with expertise, not just opinions) and how we gather information from them. Asking questions is critical here, and emphasis is again put on not jumping to solutions, even though it's so tempting!

There is also discussion about the use of questionnaires – including some guidance about where to start (at the end) and how to get the best results – keeping the questions simple, testing understanding of the questions before launch and ensuring you know exactly how you're going to use the data.

Listen on buzzsprout

https:// evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 3.1: Questionnaires used in your organization

1. Make an overview of the questionnaires and surveys your organization uses (or has used in the past) to gather information from clients, employees, or others. Indicate the purpose of each questionnaire, how many items (questions) it contains, how it is administered (do you know the response rate?) and how it was developed (e.g., home-grown, purchased from a vendor).
2. Select one questionnaire and critically evaluate the wording of its items. If you feel a question is not adequately formulated, make suggestions for improvement.

Submit a Word document with your overview, the questionnaire you have selected, and your suggestions according to the instructions of your professor/instructor.

Exercise 3.2: Gathering evidence from practitioners

1. Think about a management/business/policy decision or change project you have been involved in (or have observed closely). This decision/project should be reasonably important for your organization, involving significant resources and several or many people. Describe briefly
   - What is (was) the problem to be solved?
   - What is the proposed (or implemented) solution?
2. How would you go about acquiring evidence from experienced professionals in your organization regarding the nature of the problem or the usefulness of the solution?
3. Write down in detail
   - From whom would you seek out evidence? Why these specific individuals?
   - How would you avoid (or reduce) selection bias?
   - What method would you use to obtain this evidence?
   - What questions would you ask?

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

References

1. Squire, P. (1988). Why the 1936 Literary Digest poll failed. *Public Opinion Quarterly*. Volume 52. Number 1. 125-133.
2. Rowe, G. & Wright, G. (2001). Expert opinions in forecasting: Role of the Delphi technique. In S. J. Armstrong (Ed.), *Principles of forecasting: A handbook of researchers and practitioners*. Kluwer Academic Publishers. Boston.
3. Armstrong, S. J. (2001). Combining forecasts. *In Principles of forecasting: A handbook for researchers and practitioners*. Kluwer Academic Publishers. New York.
4. Yaniv, I., & Choshen-Hillel, S. (2011). Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making*. Volume 25. Number 5. 427–434.
5. Bauer A., Eisenbeis, R. A., Waggoner, D. F., & Zha, T. (2003). Forecast evaluation with cross-sectional data: The blue chip surveys. *Economic Review*. Volume 8. Number 2. 17-32.
6. Miller, P. R. Tip sheet—Question wording. *Duke Initiative on Survey Methodology (DISM)*. https://dism.ssri.duke.edu/survey-help/tipsheets/tipsheet-question-wording.
7. Example adapted from Survey Monkey, *Writing good survey questions.* Copyright © 1999-2020. www.surveymonkey.com/mp/writing-survey-questions.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 3.1

A

1. Incorrect > The sample size should depend on how many employees are employed and how accurate you want your evidence to be, but in general, a sample of 150 should give you a fairly reliable impression of the employees' satisfaction.
2. Correct > Employees who start their working day on Monday at 9 a.m. may differ from employees who start their working day at 10 a.m. or even later. Also, some employees with a part-time contract may have a regular day off on Monday. This means that these two groups of workers are likely to be underrepresented in the survey.
3. Correct > Administering the survey on Monday at 9 a.m. is not random and will therefore yield a sample that is not representative of all employees.

B

1. Incorrect > Some people may arrive late (or even on the second day) at the conference, so their views would be underrepresented.
2. Correct. People present on the second day are most likely representative of the conference's attendants.
3. Incorrect. People who stay until the end of a conference are likely to be more committed than people who leave early, so their views would be overrepresented.

Did I get this? 3.1

a. Yes > You may be seen by some practitioners as an authority and as a result get less critical answers.

b. Yes > You may unconsciously filter out all answers that do not confirm your own judgmen

c. No > Group conformity is the tendency to conform to the others in a group. In this example, there is no group to conform to.

d. Yes > You may encounter only practitioners who are not representative of the target audience (e.g., because you wander around only on the floor whe e your office is located)

Did I get this? 3.2

1a. No > Because participants respond anonymously to the questions, the risk of authority bias decreases.

1b. No > The Delphi method does not involve searching, weighing, or selecting information/evidence, so there is no risk of confirmation bias

1c. No > Because participants respond anonymously to the questions, the risk of group conformity decreases.

1d. Yes > Selection bias occurs when the group of practitioners that participates in a discussion is not representative of the whole population of practitioners, so selection bias can occur in the Delphi method.

2a. Yes > Because participants openly respond to the questions and provide their opinion, the risk of authority bias increases.

2b. No > In general, a face-to-face meeting does not involve searching, weighing, or selecting information/evidence, so there is less risk of confirmation bias

2c. Yes > Because participants openly respond to the questions and provide their opinion, the risk of group conformity increases.

2d. Yes > Selection bias occurs when the group of practitioners that participates in a discussion is not representative of the whole population of practitioners, so selection bias can occur in a face-to-face meeting.

3a. No > Because participants respond anonymously to the questions, the risk of authority bias decreases.

3b. No > A mobile voting system does not involve searching, weighing, or selecting information/evidence, so there is no risk of confirmation bia

3c. No > Because participants respond anonymously to the questions, the risk of group conformity decreases.

3d. Yes > Selection bias occurs when the group of practitioners that participates in a discussion is not representative of the whole population of practitioners, so selection bias can occur with a mobile voting system.

Did I get this? 3.3

1a. No > This is a double-barreled question because *roles and responsibilities* and *clear and consistent* are different things. An option would be to split the question into two or more separate questions.

1b. Yes > The words that are used in this question are clear and unambiguous, so this is an adequately formulated question.

1c. No > In this question, the term *regular* is imprecise and should be avoided. Better would be to ask, “In the past month, how many times was risk management an agenda item at team meetings?”

1d. No > This question is vague and uses abstract terms such *open environment* and *receptive to challenge*. Better would be to define these terms or illustrate them with an example

2.1. D > People will be inclined to give a socially desirable answer.

2.2. D > This question is double barreled – knowing the outcome does not equal effective feedback. Splitting the question into two separate questions would be better.

3. A and C > This question is a poorly worded question: it is double-barreled (*changes to regulations* and *how they affect me* are two separate things). In addition, it can be difficult is vague. As a result, most respondents will be inclined to answer yes to this question.

4a. This question is ambiguous because of the word *sometimes*. By omitting this word and instead using a Likert scale (Always, Very Often, Occasionally, Rarely, Never), the question could be improved.

4b. This is obviously a double-barreled question, but it also uses loaded and emotional language, which can bias respondents to provide a certain answer. Terms such as *very*, *strong*, and *excellent* should therefore be omitted or replaced with more neutral language.

4c. This is an adequately worded question.

4d. This is a double-barreled question: (1) commitment to completion, (2) accuracy, (3) timeliness, and (4) attention to detail). It also uses several abstract/ambiguous terms. Terms such as *commitment to completion* should be defined or illustrated with an example

CEBMa
center for
Evidence-Based Management

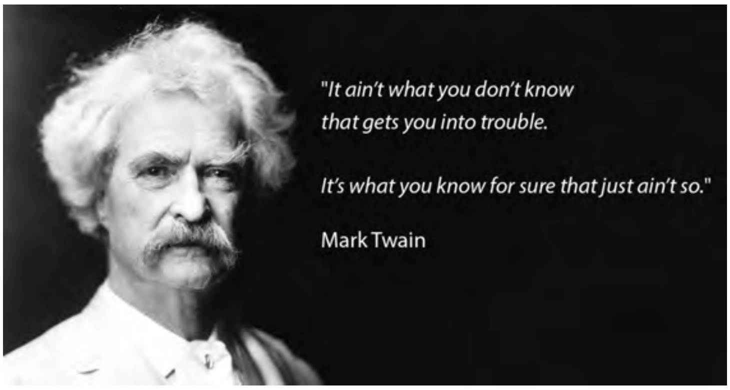
Open Learning Initiative

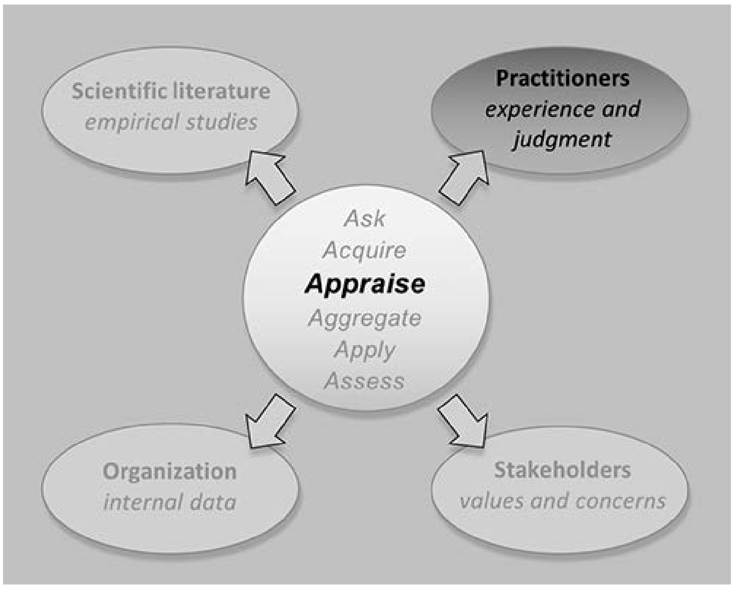
MODULE 4 | APPRAISE Evidence from Practitioners

Learning objectives:

- Determine whether professional expertise is valid and reliable.
- Identify/recognize System 1 and System 2 thinking.
- Identify common cognitive biases.
- Propose measures to overcome cognitive biases.

Introduction





Three real-life examples

Example 1

Cisco, a Silicon Valley firm, was once the darling of the new economy. Business analysts praised its customer service, perfect strategy, unique corporate culture, and charismatic CEO. In March 2000, it was the most valuable company in the world. When Cisco's stocks plummeted 80 percent the following year, the analysts reached a different conclusion: poor customer service, a vague strategy, a lame corporate culture, and a weak CEO. However, neither the strategy nor the CEO had changed. What had changed was the demand for Cisco's products – and that was through no fault of the firm. [1]

Example 2

For decades, physicians believed that a stomach ulcer was caused by lifestyle factors, such as severe stress. They drew this conclusion because they noticed that men and women with high levels of occupational or personal stress were more likely than others to develop a stomach ulcer (the assumption was that stress led to excess stomach acid, which in turn damaged the stomach's lining). Accordingly, treatment was aimed at neutralizing acid secretion through special diets and helping patients learn how to deal with stress. However, in the early 1980s, it was found that a stomach ulcer was caused by a bacterial infection that could easily be treated with an antibiotic. This meant that the causal relation that physicians have seen for decades turned out to be nonexistent.[2]

Example 3

In 1998, McKinsey – America's largest and most prestigious consulting firm – wanted to examine how the top-performing companies di fered from other firms. For this purpose, the firm surveyed and interviewe thousands of managers across the country. When the three consultants who headed the project sifted through the results, they noticed a pattern. To win in business, they concluded, companies must find and hire as many top performers as possible and then promote their most talented people aggressively. In 2001, the consultants published their insights in a book, *The War for Talent*, in which 27 companies were cited as representing best practices in the industry. Within 5 years, however, most of these companies had either disappeared or reported disastrous profitability and investment eturns. One of these companies was Enron, considered by McKinsey to be the ultimate "talent" company. It came to light that Enron's top executives had lied about its profits and had used very cleve -seeming but illegal practices to increase revenues. As a result, the company's shares plummeted in value, and in December 2001, a few months after the publication of *The War for Talent*, Enron filed the largest bankruptcy in US history, leaving tens of thousands unemployed and with worthless stock in their pensions.[3]

In organizations, evidence from practitioners is an important source of information. Unfortunately, of the available sources of evidence, professional judgment and expertise are most prone to bias. Therefore, the process of appraising professional expertise and judgment requires explicit assessment as to what extent this evidence may be biased. In this module, you will develop a better understanding of the nature of professional expertise and detect common cognitive biases that may negatively affect practitioners' (and your) judgment.

What constitutes valid and reliable professional expertise?

In general, we regard engineers, medical specialists, and management consultants as highly educated and experienced professionals. However, sometimes even highly educated and experienced professionals hold erroneous beliefs, not because they are ignorant or stupid but because their judgment is shaped by misreadings of their own personal experience. This leaves us with some fundamental questions: How trustworthy is the judgment of experienced professionals? What constitutes valid and reliable evidence from practitioners?

Professional experience is about the time a professional has spent on doing a task or activity in a certain domain. It is the foundation for acquiring ‘expertise’: the deep knowledge and skills developed in that domain. Expertise differs from intuition and personal opinion because it reflects the specialized knowledge acquired through education, training, and the in-depth practice of specialized activities. It is an essential component for determining whether a management issue really requires attention, whether the available organizational data are trustworthy, whether research findings apply, or how likely a proposed solution is to work in a particular context.

However, accumulated experience alone does not necessarily result in reliable expertise—even professionals with years of experience can do the wrong things without realizing their mistake.(4) Only under four specific circumstances does professional experience lead to valid and reliable expertise. (5) (6)

1. When the domain or activity is sufficiently narrow and specific

2. When there are opportunities for repeated practice

3. When practice leads to direct, objective feedback

4. Within a regular, predictable work environment

Learn by doing 4.1

Over a five-year period, whose professional expertise would you judge to be the most reliable? Rank order the following three professionals (most valid and reliable on top).

- An orthopedic surgeon specializing in knee surgery
- A management consultant specializing in mergers and acquisitions of hospitals
- A baker specializing in making sourdough bread

Our answer:

We would argue that a baker’s professional expertise can be considered the most valid and reliable: making sourdough bread involves a narrow, specific activity, and the baker works in a bakery, which is a regular and predictable work environment. In addition, the baker bakes multiple loaves of bread each day and so has plenty of opportunity to practice and receive direct feedback. Finally, the baker’s success or failure is very clear because they can observe the results firsthand as well as receive customer feedback.

In addition, that the orthopedic surgeon's professional expertise can be considered fairly valid and reliable: knee surgery is a fairly narrow domain and the surgeon works in a highly controlled environment (an operating room), where the results of his or her actions can be readily determined, and direct feedback is obtained when the surgeon sees the patient the next day and during follow-up visits. In addition, it tends to be clear whether the surgery was successful: relevant outcome metrics, such as increased mobility and reduced pain, can be measured in a valid and reliable manner. Finally, most orthopedic surgeons perform a specific surgical procedure several times a week. However, compared to baking bread there are more (partly unknown) factors that may affect the outcome of a surgery, so we would judge the orthopedic surgeon's expertise less valid and reliable

Finally, the management consultant's expertise cannot be considered valid and reliable. A merger or acquisition is highly complex, involving many aspects, and cannot be classified as a narrow domain or specific activity. Additionally, a management consultant specializing in mergers and acquisitions does not typically operate in a regular and predictable environment. The outcome of a merger is often influenced by numerous contextual factors, such as organizational differences, power struggles, and the economic situation, which make it hard to determine whether the outcome was the result of the consultant's actions or other factors. In addition, it is often not immediately clear whether a merger was a success or a failure: results may be difficult to determine, and what is regarded as a success by one person may be seen as a failure by another.

Learn by doing 4.2

Think of an example of a management role that operates within a relatively steady and predictable work environment and receives frequent, direct, and objective feedback. Create a detailed description in the space provided.

Did I get this? 4.1

Rank these options in terms of whose professional expertise you would judge to be more valid and reliable, over a five-year period, the most valid and reliable on top and the least valid and reliable at the bottom.

- A trader at the stock exchange specializing in gold stocks
- An HR manager specializing in talent management
- A clinical psychologist specializing in anxiety disorders
- A carpenter specializing in wooden chairs

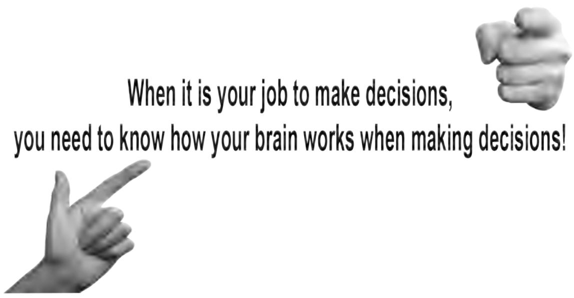
Please Note!

By a regular, predictable work environment, we mean one that is free from unforeseen organizational factors that might randomly influence the outcome of the activity. This is distinct from exposure to varying practice conditions, which can help professionals develop the ability to adjust their thinking and successfully adapt their practice to changing circumstances. In fact, intentionally seeking out practice situations with diverse conditions and reflecting on the outcomes appears to be a crucial factor in developing valid and reliable professional expertise.

Why is it so hard to develop valid and reliable professional expertise?

The organizational domain often is unfavorable to developing valid and reliable professional expertise. (7) There are, of course, exceptions. A good example is a sales agent. But even having many opportunities in which to practice, receiving direct and objective feedback, and working in a controlled, predictable environment do not guarantee that our experience is valid and our judgment sound. (8)

It was long believed that human beings base their judgment on experience, knowledge acquired through education, or other sources of information. However, in the past 50 years, an abundance of research has repeatedly demonstrated that our judgment is highly susceptible to systematic errors – cognitive and information-processing limits make us prone to biases that have negative effects on the quality of the decisions we make. (9)



Notably, four Nobel prizes were awarded to researchers Herbert Simon, Daniel Kahneman, Robert Shiller, and Richard Thaler, whose scientific work demonstrates that human judgment systematically deviates f om rationality. These systematic errors are the result of the way our brain is wired: we are predisposed to see order and causal relations in the world, we are overly optimistic, we are overly confident, and we process information in a way that confirms our existing beliefs, expectations, and assumptions. As a result, cognitive biases are more influential than you might think, and they are the secret author of many professional judgments. (10)

Two modes of thinking

We make judgments every day – whether we can trust a person, whether we should do something (or not), which route to take, how to respond to someone's question – the list is endless. If we carefully considered and analyzed every possible outcome of these judgments, we would never get anything done! Thankfully, our mind makes things easier by using efficient thinking strategies known as *heuristics*.

Heuristics

A heuristic is a mental shortcut that helps us make judgments quickly without having to spend a lot of time researching and analyzing information. They allow subconscious mental processes to make up for the lack of information and lead us to routine decisions that are often correct. Most of the time, we apply heuristics below the radar of our conscious awareness, and so we are often oblivious to the impact of heuristics on our judgments. They are learned or hardwired into our brain by evolutionary processes.

Following are several types of heuristic and examples of the logic behind them:

Example

- **Authority heuristic:** "This man has a degree in botany from Harvard, so if he says this flower is no a dandelion but a salsify, I had better trust his judgment."
- **Wisdom of the crowd:** "If this many positive reviews have been written about the hotel, and if this many people recommend it, then it must be good."
- **Representativeness heuristic:** "This woman loves to listen to New Age music and faithfully reads her horoscope each day, so she is more likely to be yoga teacher than a bank teller."
- **Halo effect:** "This candidate looks very professional—he wears a nice Italian suit and has a warm, confident smile, so he s probably a great manager."
- **Educated guess:** "This house has a garden, so I guess it's more expensive than the one with only a balcony."
- **Familiarity heuristic:** "When I'm on vacation and have to buy groceries, I always buy brands that I recognize."

Heuristics clearly influence our decisions, but we do not always form our judgment or make decisions by using heuristics. Sometimes, we take the time to make deliberate, mindful decisions based on a careful weighing of all the information available. In his book *Thinking, Fast and Slow*, Nobel prize–winning psychologist Daniel Kahneman refers to these two modes of processing as System 1 and System 2 thinking. (10)

System 1 and System 2 thinking

System 2 is slow, effortful, deliberate, and rational. It is the slow, effortful reasoning system that draws heavily on our cognitive resources and requires attention and concentration. In fact, all thinking that demands mental effort tends to be classified as System 2. In contrast, System 1 is fast, intuitive, associative, and emotional. It is the fast, effortless thinking system that operates automatically with little voluntary control and that uses intuition or heuristics to make decisions fast.

In daily life, System 1 is our dominant mode of thinking, because this way of information processing is necessary for survival. When we are driving, System 1 automatically makes us hit the brakes when the brakelights of the car in front of us come on – we don't have time for System 2 to process all information and figu e out whether the other driver sees a dangerous situation ahead or is just an extremely nervous person with limited driving experience.

In addition, when a number of students suddenly jump up and leave the room, System 1 triggers other students to follow suit. From an evolutionary perspective, this behavior makes sense: there could be a dangerous situation, such as a fi e, that made the students decide to leave the room. After carefully processing, weighing, and judging all information System 2 may conclude that their decision was wrong (there is no fi e—it's just the air conditioning leaking a foul-smelling water vapor), but in life-and-death situations, there is often no time for careful judgment.

System 1 and its heuristics can speed up our judgment and decision-making process, but it also can introduce *serious cognitive biases that impair the quality of the decisions we make*.

Learn by doing 4.3

Try to answer the following question within 10 seconds.

- A bat and a ball cost $1.10 in total.
- The bat costs $1 more than the ball.
- How much does the ball cost?

The interaction between System 1 and System 2 defines how we think. However, we rely on System 2 much less than we think we do. Often, when we presume we are thinking rationally and systematically – when we think we have a sound reason for our judgment – in effect our judgment is tainted by the heuristics and biases that are dictated by our System 1. So, why don't we use System 2 more? According to Daniel Kahneman, it is because using System 2 is hard work. "The law of least effort applies. People are reluctant – some more than others, there are large individual differences. But thinking is hard, and it's also slow. And because System 1 thinking is usually so efficient, and usually so successful, we have very little reason to work very hard mentally. (11)

Learn by doing 4.4

Read the following four scenarios and determine whether the decision was based on System 1 or System 2 thinking.

1. A consultant gives a presentation that shines with professionalism: the slides and graphics are well designed, and the model he recommends has a logical structure and appears to make sense. The organization's senior managers therefore decide to hire the consultant and adopt the model.

2. A large academic hospital loses revenues because patients don't show up for their scheduled appointment. The IT director therefore suggests to implement a text-messaging system that sends a reminder to patients 24 hours before their appointment. Most physicians think this is a great idea that will certainly reduce the number of no-shows. The hospital's administrator, however, decides that he first wants to find out what the experiences are of hospitals that have already implemented the system.

3. A well-known, prestigious international consulting firm recommends introducing performance bonuses for executives and senior-level employees in order to boost the company's performance. The CEO decides that she first wants to know what the research literature suggests about the benefit (effectiveness) and costs (possible negative effects) of performance bonuses.
4. A not-for-profit organization has decided to hi e a consulting firm to support the development of new compliance management system. The CEO suggests inviting one well known and two lesser-known consulting firms to tender a proposal under the condition that the proposal is anonymized (i.e., the name of the firm must be left out of the proposal).

Did I get this 4.2

Read the following scenario and determine whether the decision was based on System 1 or System 2 thinking.

The CEO of a large international accounting firm decides to introduce Agile working. He first learned about this method at an international seminar and noticed that, in the past year, many banking firms have introduced agile working. He feels that if so many companies have adopted this new way of working, it must be effective.

Cognitive biases

Cognitive biases are errors in thinking that affect how we make decisions. They stem directly from our System 1. There are many different types of cognitive biases. For example, Wikipedia provides a list of more than 100 different biases, some of them labeled with exotic names such as the *cheerleader effect* and the *gambler's fallacy*. We describe in the following pages three cognitive biases most relevant to management as well as other common cognitive biases.

1. **Patternicity and the illusion of causality**
2. **Confirmation bias**
3. **Group conformity**
4. **Other common cognitive biases**

As explained at the beginning of this module, the less the four conditions for gaining valid and reliable experience are met, the more likely a professional's expertise will be affected by cognitive biases.

Patternicity and the illusion of causality

Our System 1 is predisposed to seek order and causal relations in the world. Consequently, we tend to see patterns all around us, some of which are meaningful and others that are but meaningless noise. This cognitive bias is often referred to as *patternicity or illusion of causality*. We seek patterns and assume causal relations by connecting the dots: *A* appears connected to *B*, so we assume there is a causal link.

Learn by doing 4.5

1. **Describe what you see in the picture below.**

Our answer:

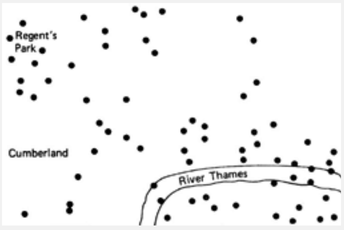
Most people see a cow. Our System 1 is very good at recognizing images of animals, faces, or objects in blurry pictures. However, our System 1 can easily be fooled: it also recognizes images when they are not there. That's why we sometimes see images of animals or UFOs in clouds or the face of Jesus on the surface of a grilled cheese sandwich.

2. This cinnamon bun (a little pastry) was first discovered at the Bongo Java coffee shop in Nashville. The coffee shop had the bun on display for about ten years, until Christmas day in 2007, when the bun was stolen. The thief later tried to sell the bun on eBay.
Can you guess why?

Learn by doing 4.6

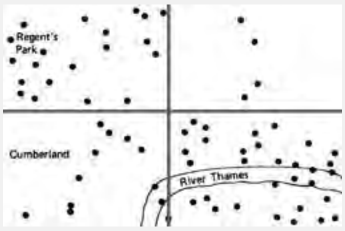
During World War II, London was struck by the Germans with V1 rockets. All points of impact were meticulously plotted on a map. When examining the map, the British Secret Service recognized a pattern: the locations at which the V1 rockets landed seemed to cluster around the government buildings at the River Thames, so they assumed that these areas were being specifically targeted. They therefore suggested moving important civil services to safer parts of the city.

Please have a look at the picture below. Do you see a pattern?

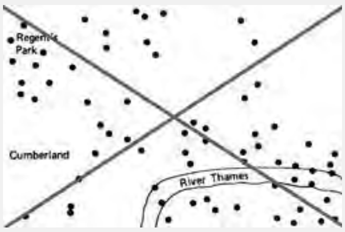


Our answer:

It looks as if the area around the River Thames has more impact points, suggesting a pattern. This pattern becomes even clearer when we divide the map into four quadrants.



Now, have a look at the picture below, where the cross dividing the four quadrants is replaced with a negative cross. *The pattern seems to have disappeared!*



After the war, statistical analysis showed that the distribution of impact points was totally random. The apparent pattern that the people from the secret service had perceived was just a coincidence. (12)

Sometimes *A* really is connected to *B*, but sometimes it is not. For instance, when someone eats a berry *(A)* and then gets sick *(B)*, she assumes the berry was poisonous. However, the baseball player who forgets to shave before the game *(A)* and hits his first home run *(B)* may falsely assume that not shaving before the game made him hit a home run. This is how people become superstitious. When a pattern or association is real, we learn something valuable from the environment, so we can make predictions that help us to survive. This process is also known as *association learning* and is fundamental to all human behavior. Unfortunately, our System 1 is not good at distinguishing false and real patterns and causal relations. In fact, human brains are inclined to believe that a causal relation is real until proven otherwise.

Did patternicity have, at some point in our evolution, added value?

Cognitive biases and heuristics are widely studied by biologists, psychologists, and neuroscientists. Most of them think that patternicity has an evolutionary origin and helped our ancestors to survive.

Imagine you are walking on the plains of Africa three million years ago. Suddenly, you hear a rustle in the bushes. You quickly look toward the sound and see a pattern of stripes. It could be just the wind and some colored leaves, but it could also be a tiger causing the rustle. If you assume it is a tiger – believing the pattern is real – and run from it, but it turns out to be just the wind, no harm done. However, if you assume it is just the wind – believing the pattern is not real – but it turns out to be a tiger, you are lunch and your life stops there.

Assessing the difference between a real and an unreal pattern or causal relation is very hard, especially in split-second situations, but in prehistoric life, it could mean the difference between life and death. As a result, our brain's default position is to assume that all patterns and perceived causal relations are real until proven otherwise. (13)

Did I get this 4.3

Give an example of how patternicity could occur in a business or workplace situation.

Our answer:

An insurance company introduces a decentralized organizational structure in which autonomous business units make their operational decisions and are fully responsible for their own profit and loss. However, two years later, the company's overhead is too high and its profit margin has decreased by 50 percent. The company's executive board jumps on the idea that a recent change is at fault: autonomous business units were introduced two years ago, and now profit margins have declined. They therefore decide to establish a new financial control system and reduce the autonomy of the business units.

Confirmation bias

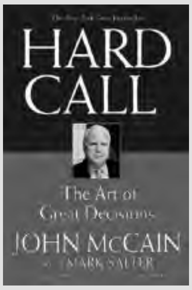
Because our System 1 dominates our thinking, we are predisposed to confirm our existing beliefs. By selectively sea ching for and interpreting information in a supporting fashion while ignoring information to the contrary, we reinforce our existing beliefs. In other words, we see what we want to see. This phenomenon is known as *confirmation bias* and is one of the most important biases.

Example

1. Confirmation bias: Astrology

You may believe that astrology actually works. As a result of confirmation bias, you'll remember only those instances when the prediction in the astrology column came true and forget the majority of the cases when the prediction was wrong. As a result, you will continue to believe astrology has some base in reality.

2. Confirmation bias: Political beliefs



Confirmation bias is particularly dominant when it comes to ideology or political beliefs. This is why it is easy to predict which news channels conservatives and liberals choose to watch and which books they like to read.

During the 2008 US presidential election, researchers analyzed the buying behavior of people on Amazon. (14) They found that people who bought Barack Obama's autobiography tended to support Obama in the election. In contrast, people supporting the rival candidate, John McCain, tended to buy books painting Obama in a negative light. Both groups appear to prefer confirmatory information

Did I get this 4.4

Confirmation bias is often present in the context of employee selection. An abundance of research shows that most recruiters and HR managers form a distinct opinion about a candidate on the basis of his or her letter of application.

Which of the following would be the BEST choice to avoid confirmation bias when hiring?

1) Letting the applicant instead submit a short video presentation of himself or herself
2) Letting an external recruitment firm do the first screening
3) Blinding information in the letter such as age, gender, and ethnicity
4) Leaving out the applicant's photo

Example

Confirmation bias: Orchestra auditions for women

Traditionally, women have been underrepresented in American and European orchestras. Renowned conductors have asserted that female musicians have 'smaller techniques,' are more temperamental, and are simply unsuitable for orchestras, and some European orchestras do not hire women at all," according to Princeton University All News. (15) Researchers Cecilia Rouse (Princeton University) and Claudia Goldin (Harvard University) examined the value of blind auditions (in which the performer is hidden from judges behind a screen) in the hiring of women in symphony orchestras. (15)

They analyzed audition records and rosters from the 1940s through the 1990s and found that blind auditions increased by 50 percent the probability that a woman would advance from preliminary rounds and by 25 percent the probability that she would ultimately be hired. Since the 1970s, blind auditions have been adopted by most American orchestras, resulting in "a significant impact on the face of symphony orchestras. About 10 percent of orchestra members were female around 1970, compared to about 35 percent in the mid-1990s. Rouse and Goldin attribute about 30 percent of this gain to the advent of blind auditions."(16)

Did I get this 4.5

Read the following scenario.

Ricardo Semler
at TED Talk (2014)
Video link

Starting in the 1980s with details such as dress codes and parking spaces, Brazilian business leader Ricardo Semler gradually created an organization where the workers recruited their own bosses, managers set their own salaries, and anyone could attend any meeting (though meetings were kept to a minimum). By the early 2000s, Semler had more or less made himself redundant, but under his ownership, his company Semco – a conglomerate with a diverse business portfolio from machinery to environmental consulting – experienced very high levels of growth and profits, with very low staff turnover. (17)

Your CEO is very enthusiastic about this new, revolutionary leadership model and considers implementing it in the organization (an international bank). He suggests paying a visit to Semler's company to see if this new model is indeed as successful as claimed. However, if your CEO would visit Semler's company, he might hear only success stories that will confirm his prior (positive) beliefs regarding the leadership model.

What would you advise your CEO to prevent falling prone to confirmation bias?

1) Cancel the visit and instead ask Semler's company to send relevant documents.
2) Ask many critical questions during the visit.
3) Also pay a visit to a company where the implementation of the model has failed.
4) Ask the managers for hard, objective evidence that the model is a success.

In his book *You Are Not So Smart*, (18) David McRaney explains that decades of research have placed confirmation bias at the top of all cognitive biases and among the most important mental pitfalls. For this reason, journalists who want to tell an objective story must actively search for evidence contradicting their initial beliefs. Without confirmation bias, conspiracy theories would fall apart. Did the Americans really put a man on the moon? McRaney argues that if you are looking for evidence they didn’t, you will find it. In this age of the Internet, people are bombarded with information. However, our System 1 tends to filter out information that does not confirm our existing beliefs. In making an evidence-based judgment, we therefore need to actively search for evidence that challenges our judgment.

Example

Confirmation bias: Six Sigma

If you believe that techniques such as Six Sigma are likely to make business processes more efficient, it won't be difficult to find evidence confirming your belief. A wide range of companies claim to have successfully implemented Six Sigma, among them Motorola, General Electric, Amazon, Ford, and even the United States Army (19) – a sufficient number to make your System 1 automatically jump to the conclusion that your belief is right. However, if you were to suspend your judgment and actively search for evidence to challenge your belief, you would find several examples of companies where the implementation of Six Sigma was less successful.

For instance, 3M, one of the world’s most innovative companies, lost its innovative edge when it began using Six Sigma to try to improve its operational efficiency. James McNerney, the CEO named in 2000, introduced Six Sigma as soon as he took the helm of the firm. But when he applied Six Sigma to 3M s research and development processes, it led to a dramatic fall-off in the number of innovative products developed by the company during those years. (20) In fact, *Fortune* reported in 2006 that 91 percent of the large corporations that had implemented Six Sigma had fallen behind the growth rate of the S&P 500, blaming this poor result on a significant decline in innovation at these firms (21)

Suspending your judgment in order to actively search for contradicting evidence is an effective way to prevent confirmation bias. In this particular case, we would probably conclude that many examples do indeed indicate that Six Sigma may have benefits in terms of cost efficiency and quality control but that there is also evidence to suggest that it may seriously hamper innovation.

Group conformity

Group conformity is the tendency to conform to the others in a group, even if doing so goes against your own judgment. Human beings are very social creatures and are hyperaware of what people around us think. Therefore, our System 1 is strongly inclined to conform to the group: we strive for consensus and avoid confrontations even when we don't agree with what people are saying.

The following video shows a humorous example of group conformity.

Another, more disturbing example of how people are naturally inclined to conform to the group is the Asch experiment, (22) a study that was first conducted in the 1950s but that has been repeated numerous times with exactly the same outcome.

Example

https://youtu.be/jw2Y3dUm7Mc

Example

https://youtu.be/AyEEGfQ2IuY

In a professional context, conformity bias can have devastating effects on the outcome of a decision-making process. In fact, there are numerous examples of conformity bias affecting a professional judgment made by clever people. The most (in)famous may be the disaster in 1986, when the NASA Space Shuttle *Challenger* broke apart 73 seconds into its flight, leading to the deaths of its seven crew members. In the corporate world, similar examples exist, such as the collapse of Swiss Air, (23) Enron, (24) and the global financial crisis of 2008

We can see another example of conformity bias in the context of management and organizations in popular management techniques. When you look at a chronological overview of popular techniques, it is hard to get a sense of scientific p ogress: management by objectives, business process re-engineering, total quality management, learning organizations, knowledge management, lean management, Six Sigma, talent management, employee engagement, Agile. Though many of these techniques once enjoyed the enthusiastic support of managers and consultants, all but the most recent have fallen from favor, replaced by the new flavor of the month. As such, the ebb and flow of popular management techniques is similar to that of a fashion cycle. (25) (26)

The impact of group conformity in the domain of management is strikingly worded by Geoffrey Colvin: "And there we see the power of any big managerial idea. It may be smart, like quality, or stupid, like conglomeration. Either way, if everybody's doing it, the pressure to do it too is immense. If it turns out to be smart, great. If it turns out to be stupid, well, you were in good company and most likely ended up no worse off than your competitors." (27)

The message here is clear: practitioners, like all human beings, are social creatures who are very much aware of what other practitioners think. As a rule, they don't want to be seen as the person who rocks the boat, so in many cases, they will conform to the group.

How was group conformity, at some point in our evolution, helpful for survival?

Simply stated, we must conform to survive. We are all autonomous individuals seeking our own personal objectives. However, we are also members of groups. Since the earliest days of mankind, human beings have been living in groups, such as bands of hunters, clans, political parties, and tribal communities. Through evolutionary selection, human beings have learned that forming alliances – for instance, to hunt and gather food – provided beneficial outcomes. Disrupting the alliance would make it ha der to reap the benefits achievable through group effort. As Darwin explained in *On the Origin of Species*, “A tribe including many members who were always ready to aid one another, and to sacrifice themselves for the common good, would b victorious over most other tribes; and this would be natural selection.” (28) Another reason for group conformity is the need for social acceptance. We are very social creatures. Joining a group satisfies our need to belong and defines our sense of self and social identity. As deviant behavior (or opinions) may lead to ejection by the group, we thus benefit f om (temporarily) disregarding our own preferences and personal objectives simply to be perceived as one of the group.

It was long thought that groups are most likely to experience conformity bias and groupthink when they are highly cohesive, are insulated, lack impartial leadership, lack methodological decision-making procedures, and experience conditions of high stress. The scientific evidence, however , shows a mixed body of results. (29) Group homogeneity (of members' social backgrounds and ideology), group insulation, poor decision-making processes (limited search and appraisal of information, incomplete survey of alternatives, failure to consider alternatives, confirmation bias in processing information), and lack of impartial leadership are generally supported as antecedents of groupthink. (30)

Did I get this 4.6

Read the following scenario.

In 2015, the business magazine Fortune looked at the number of Fortune 500 companies with entirely male boards. Their disappointing conclusion was that the number was in the double-digits: 23. In addition, most board members of these companies were white, had an MBA degree from a top business school, and shared a similar political orientation.

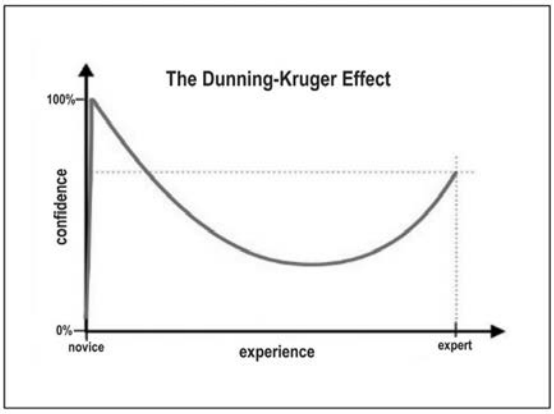
Why could this group homogeneity affect the quality of the decisions made by these boards?

Other common cognitive biases

The effect of cognitive biases on human judgment is one of the most widely studied topics in the field of psychology. As a result, an overwhelming number of scientific publications are available. As previously mentioned, Wikipedia lists more than 100 different biases. One could argue that all cognitive biases are relevant to the domain of management. However, apart from the patternicity, confirmation bias, and group conformity, we consider the following biases particularly relevant:

- **Availability bias** – *The tendency (of our System 1) to rely on examples that spontaneously come to mind when evaluating a specific topic, method, or decision.* As a result, we are more likely to believe something is true, effective, or commonplace if we know an example of it, and we are less likely to believe in something that we've never seen or heard before. However, the examples that are "available" in our memory are largely determined by how recent the examples are and how unusual or emotionally charged they may be. For example, a manager who has just read an article on the benefits of lean Six Sigma in a popular magazine will more likely suggest (or accept) this as a solution for a managerial problem than a model he or she has never heard of.
- **Authority bias** – *The tendency to overvalue the opinion of someone who is seen as an authority.* A person's status and credentials greatly influence the way we perceive that person's message. As a result, we tend to be less critical when an "authority" makes a claim or a suggestion. For example, consultants from a large, renowned international consulting firm tend to get less critical questions when making a recommendation than does a consultant from a small, relatively unknown local firm. The same counts for the judgment of the so-called HIPPO—the highest-paid person's opinion—a term coined by Avinash Kaushik. (31) In meetings, HIPPOs are often deemed to be more valuable and important than the judgment of a practitioner on a lower pay grade.
- **Outcome bias** – *The tendency to evaluate the quality of a decision (intervention, method) on the basis of its outcome.* This bias occurs even when the outcome is determined by chance. For example, when Steve Jobs was fi ed from Apple in 1984 over a disagreement regarding how to save the company during a period of declining sales for the Apple II computer, he was publicly regarded as a bad leader who had seriously damaged the organization with his emotional swings and immature behavior. However, after his return in 1996 and Apple's success with the iPod and the iPhone, he was regarded as one of the most influential and innovative leaders of his time
- **Overconfidence bias** – *The tendency to have an unwarranted faith in one's own knowledge, judgments, cognitive abilities, or skills.* Research has demonstrated that 93 percent of American drivers rate themselves as better than the median. (32) In a similar survey, 87 percent of MBA students at Stanford University rated their academic performance as above the median. (33) Professionals, such as physicians, economists, lawyers, and managers, are by no means immune to this tendency: research has demonstrated they all tend to overestimate their knowledge and ability. Physicians, for example, frequently underestimate the proportion of negative outcomes among clients in their caseloads. (34)

A special type of overconfidence bias is the Dunning-Kruger effect, named after David Dunning and Justin Kruger of Cornell University. In their seminal paper "*Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Leads to Inflated Self-Assessments*," they show how incompetent people fail to adequately assess their level of competence and thus consider themselves much more competent than everyone else. (35)



Learn by doing 4.7

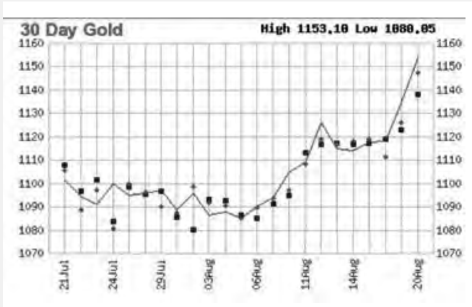
Please read the following scenario and answer the questions below.

A common cliché in sport is that you are only as good as your last game. But does this also apply to surgeons? Imagine you have a medical condition for which you need minor surgery. In the hospital, you must choose between two surgeons, surgeon A and surgeon B. Surgeon A's last operation had a dramatic outcome. During the operation, serious complications occurred and the patient suffered permanent damage. Surgeon B's last operation, however, was very successful, and the patient recovered faster than expected.

1. **Based on this outcome, most people will be inclined to choose surgeon B. Which of the following biases is influencing this choice?**
   a. Group conformity
   b. Outcome bias
   c. Confirmation bias
   d. Authority bias
   e. Availability bias
   f. Patternicity/illusion of causality
2. **Explain why the outcome of a surgeon's last operation is a bad indicator for his or her professional expertise.**
3. **Formulate at least two questions that you would like to have answered to judge the surgeons' professional expertise.**

Learn by doing 4.8

Please read the following scenario.



One of your best friends is a trader on the stock exchange. He enthusiastically tells you he has analyzed a large number of financial and economic data and that he has discovered an interesting phenomenon:

"The position of the Dow Jones index multiplied by the price of oil is two days ahead of the gold price!" In other words, if both the Dow Jones and the oil price go up, the price of gold will rise within two days.

1. **Given this scenario, which of the following biases is more likely to have distorted your friend's analysis?**
   a. Patternicity/illusion of causality
   b. Confirmation bias
   c. Group conformity
   d. Availability bias
   e. Authority bias
   f. Outcome bias
   g. Overconfidence bias
2. **Formulate a question that you would like to have answered to judge the reliability of the phenomenon your friend has discovered.**

Did I get this 4.7

A 2006 study (36) *found that 74 percent of the 300 professional fund managers surveyed believed that they had delivered above-average job performance. Only 26 percent of the fund managers viewed themselves as equal to or below average. Clearly, only 50 percent of the sample can be above average.*

Which of the following has likely affected the judgment of the majority of the fund managers surveyed?

a. Patternicity/illusion of causality
b. Confirmation bias
c. Group conformity
d. Availability bias
e. Authority bias
f. Outcome bias
g. Overconfidence bias

Did I get this 4.8

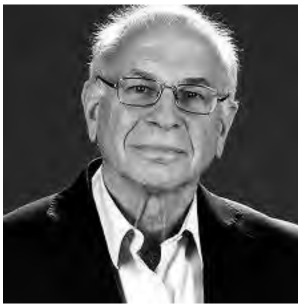
From 1997 to 2006, US housing prices rose about 85 percent. This rate of increase was five times the historical rate of 1.4 percent a year, making it the biggest rise in US history. Many people, including bankers and investors, strongly believed that the prices would continue to increase by about 10 percent a year, making it a safe investment. Even Federal Reserve Chairman Ben Bernanke claimed housing prices were solid and would never drop. Then came the crash of 2008, and within a few months, average US housing prices declined by more than 20 percent from their mid-2006 peak. As a result, $5 trillion in real-estate value, pension money, and savings disappeared, and 6 million people lost their homes. The first signs of decline, however, were already showing in 2006. Nevertheless, most financial experts did not see the crash coming.

Which of the following biases are demonstrated in this scenario?

a. Patternicity/illusion of causality
b. Confirmation bias
c. Group conformity
d. Availability bias
e. Authority bias
f. Outcome bias
g. Overconfidence bias

How can you avoid or reduce bias?

Because of the dominance of our System 1, human judgment is highly susceptible to systematic errors. The fundamental question is *How can we make professional judgments without falling prey to our biases*? Awareness that decisions can be biased is an important first step. Still, being aware that human judgment suffers from cognitive biases does not prevent them from occurring.



Even Daniel Kahneman, the world's leading authority on this subject, stated: "I've been studying this stuff for 45 years, and I'm no better than when I started. I make extreme predictions. I'm overconfident. I fall for every one of the biases." Thus, you may accept that you have biases, but you cannot eliminate them in yourself. However, as Kahneman also points out, "There is reason for hope when we move from the individual to the collective, from the decision maker to the decision-making process, and from the executive to the organization." (37)

As researchers have documented in the realm of management, the fact that individuals are not aware of their own biases does not mean that biases can't be neutralized – or at least reduced – at the organizational level. As an evidence-based professional, you can add tremendous value to your own and others' decision-making by acquiring, appraising, and applying multiple sources of evidence to the judgments you and others make. In addition, you can take some practical measures to prevent cognitive biases from clouding your own judgment (or that of other practitioners).

1. Consider multiple options

The scientific literature suggests that considering multiple solutions tends to lead to better judgment than fixating on yes/no or either/or choices. When we consider only one solution, we tend to ignore evidence contradicting its expected results, and as a result, we are prone to confirmation bias. Conversely, considering two or more options leads to gathering more information regarding expected differences in results and thus leads to a more informed judgment.

2. Get the evidence before forming an opinion, not after

As Sherlock Holmes stated, "It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts." (38) If you nevertheless have a strong opinion or preference (e.g., regarding a specific solution, intervention, or management model), it may be helpful to make your beliefs explicit, for instance, by writing it down. This way, you remind yourself that you may be biased against alternative options and take that possible bias into account when making a judgment.

3. Blind assessment

An effective method to prevent confirmation bias, halo effect, and authority bias is *blinding*. For example, hiding all information in a person's curriculum vitae and application letter that could induce bias allows making a more objective assessment of a candidate's qualities. The same is true for the source of the evidence when evaluating its trustworthiness, such as the title of the journal where a scientific study was published, the name of the consulting firm that provided the data, or the seniority or status of the professional who provided his or her judgment. After all, we are interested only in the evidence itself, not the status or authority of its source.

4. Falsify views and judgments

Actively looking for evidence that contradicts your (or your colleagues'/ clients') beliefs and opinions can lead to a more objective and balanced judgment. In addition, it may help to actively seek out people with contradictory beliefs and judgments. Our System 1 strongly favors our own ideas, so we need others to provide balance; however uncomfortable, it will pay off long term. Herbert Simon, a founder of modern organizational science, artificial intelligence, and robotics, recognized the importance of decision processes for overcoming bias. He championed a practice still used today at Carnegie Mellon University in evaluating faculty for promotion and tenure. The discussion of all faculty cases begins with a presentation by two faculty members, one chosen to develop the pro case, why the candidate should be promoted or tenured, and another chosen to give the con case, why tenure or promotion should be denied. After the pro and con have each presented their respective cases, the assembled faculty ask them questions first "with hats on" (answering the question from the pro or con perspective) and then with "hats off" (their own personal views of the case). The result is a thoughtful, rich discussion of the candidate's case.

5. Seek disagreement

Alfred Sloan, former president of General Motors, strongly believed decisions should not be made until someone had brought forward why the "preferred" option might be wrong. "If we are all in agreement on the decision—then I propose further discussion of this matter until our next meeting to give ourselves time to develop disagreement and perhaps gain some understanding of what the decision is all about." (39) Encouraging people to disagree and to be as open as possible may help prevent groupthink and authority bias from occurring. If necessary, you could assign someone to play devil's advocate.

6. Playing devil's advocate

A devil's advocate is a person who expresses an opinion that disagrees with the prevailing point of view for the sake of debate or to explore the thought further. By taking an opposite view (which he or she may not actually hold) and playing the devil's advocate role, this person seeks to engage others in an argumentative discussion process. The purpose of this process is to determine the validity of the original point of view and identify biases in its argumentation in order to increase the quality of the decision-making process. (40) When applied to a project team, the devil's advocate should be a different person for each meeting.

7. Install a red team

The concept of a red team was originally developed by the US Army during the Cold War, using a team of officers taking a Soviet (" ed") perspective to penetrate the US defensives. Nowadays, red teams are used by companies – for example, IBM, SAIC, Microsoft, and the CIA – to challenge assumptions, unearth preconceived notions, and identify symptoms of bias (especially confirmation bias and groupthink) that could affect professional judgment.

Learn by doing 4.9

Consider the following scenario:

Your CEO is very enthusiastic about a new IT system that is said to provide valid and reliable information regarding the organization's performance on relevant outcome metrics. She suggests paying a visit to a company that has successfully implemented this new system to learn from its experiences.

Which of the following would be the best advice you could give your CEO to avoid bias? Select Yes or No for each option.

a. Bring along someone who can play devil's advocate and take an opposite view.
b. Install a red team to challenge your CEO's assumptions.
c. Visit a company where the implementation of the new IT system was not successful (falsification)
d. Perform a blind assessment of the new IT system.

Did I get this 4.9

Please read the following scenario and answer the questions below:

Senior executives in a large international investment bank are concerned with the number of transactional errors made in processing trades. They strongly believe that a significant proportion of these errors are preventable and feel that this is partly due to a lack of risk awareness among employees in the back office. They decide to discuss their view with some middle managers to see whether they agree.

1. **Their decision to discuss their view with some middle managers to see whether they agree is a questionable decision because it leaves them prone to which bias?**
   a. Group conformity
   b. Confirmation bias
   c. Overconfidence bias
2. **What should the senior executives do instead to avoid bias? Select Yes or No for each option.**
   a. Install a red team
   b. Play devil's advocate
   c. Seek disagreement
3. **What other things could the senior executives do to avoid bias?**

Did I get this 4.10

If you had to compose a team that would advise you in important decisions, what elements would you take into account? Name at least three.

Many students wonder: **Can you debias people (or yourself)?**

As stated earlier, being aware that your judgment suffers from cognitive biases does not prevent them from occurring. Many scholars therefore doubt that lone individuals can "debias" themselves. (41) Some scholars, however, suggest that everyday reasoning can be improved – to a certain extent – through experience and education. (42) In his chapter "Debiasing," Richard Larrick provides an overview of what personal strategies we as individuals can apply (43):

- **Motivational strategies:** There is little empirical evidence that incentives improve decision-making. A more viable approach is to hold people accountable for their decisions. Research has indicated that, under certain circumstances, accountability leads to greater effort and use of information, which may lead to improved decision-making. However, accountability has its own problems: if we know our audience's preference for a specific decision-making process, we may be biased toward that method.
- **Cognitive strategies:** Teaching people to consider the opposite or to ask themselves, *What are possible reasons that my initial judgment might be wrong?*, has been shown to be effective at reducing overconfidence, hindsight bias, and anchoring effects. Training in specific thinking rules, logical principles, and decision rules has also shown some effect, but this type of training needs further study.
- **Technological strategies:** Decision support systems have the potential to improve individual decision-making but much more research is needed in this area.

Critically appraising evidence from practitioners

This module may leave you with a sense of disappointment. It shows that human judgment is inherently flawed – we see patterns and causal relations where they don't exist, we selectively search for and interpret evidence in a way that confirms our prior beliefs, we conform to others in a group (even when we don't agree), we overvalue the opinion of authorities, we are overconfident, and so on. And the e is not much we can do about it. You may consequently conclude that evidence from practitioners is always unreliable and thus should be ignored. This conclusion, however, would be incorrect.

As stated at the beginning of this module, professional expertise/judgment is an important source of evidence, especially as it can connect evidence from other sources (e.g., scientific research findings and organizational data) to the specific organizational context. In addition, professional expertise is the most widely used source of evidence for the simple reason that it is easily obtainable.

So, the message here is, ***Don't throw away the baby with the bathwater!*** Practitioner judgment/expertise can, under specific conditions, be a valid and reliable source of evidence provided that we critically appraise its strength/trustworthiness before we apply it to the decision-making process.

To determine whether evidence from practitioners is valid and reliable, four questions need to be answered.

Question 1: Are the four criteria for valid and reliable expertise met?

Especially when it comes to questions that assume a certain level of professional expertise on the matter (e.g., "Will this solution solve the problem?" or "What are the possible downsides or unintended consequences of this solution?"), it is important that the three criteria for valid and reliable expertise are met:

1. A domain or activity that is sufficiently narrow and specific: Is the area of practice or task domain in which the practitioner gained experience not too broad or too generic?
2. **Numerous opportunities to practice:** Does the practitioner have extensive experience with the matter (problem, preferred solution)?
3. **Direct, objective feedback:** Was the practitioner able to evaluate the outcome, and if so, was direct, objective feedback available?
4. **A regular, predictable work environment:** Can the organizational context in which the practitioner gained his or her experience with the matter be regarded as regular and predictable?

As explained, the less the four conditions for gaining valid and reliable experience are met, the more likely a professional's expertise will be affected by cognitive biases!

Question 2: Has an effort been made to reduce bias?

As discussed earlier, we can take some practical measures to reduce or prevent cognitive biases from affecting our judgment, especially in the context of group judgment. Relevant appraisal questions you can ask are

- Were multiple options considered?
- Was the available evidence assessed blind from information that could induce bias?
- Was an attempt made to falsify views and judgments (e.g., by actively seeking for contradictory evidence)?
- Was an attempt made to actively seek disagreement (e.g., from other practitioners)?
- Was an opposite view brought into the judgment process (e.g., a devil's advocate)?

Question 3: Could there be bias?

Even when we have made a serious effort made to reduce bias, it could still affect people's judgment. Cognitive biases are, by definition, implicit, meaning that bias is not a deliberate process but something that unwittingly sneaks into someone's judgment. For this reason, you should always check whether – and to what extent – cognitive bias, such as social desirability bias, conformation bias, availability bias, or group conformity, may have affected a practitioner's judgment.

Question 4: Was the evidence acquired in a valid and reliable way?

The strength/trustworthiness of evidence from practitioners is determined not only by factors such as experience, feedback, and susceptibility to bias but also by the way in which the evidence was obtained. Therefore, when critically appraising evidence from practitioners, we should also account for the method used to acquire the evidence. As discussed in Module 3, ACQUIRE Evidence from Practitioners, wandering through the workplace and asking practitioners their opinion on an assumed problem or preferred solution is a method more prone to bias than conducting a survey with a large random sample or a Delphi panel.

In addition, a major concern when obtaining evidence from practitioners is selection bias. This occurs when your selection of practitioners leads to an outcome that is different from the one you would have expected had you enrolled the entire target audience.

Finally, we must consider the wording of the questions when acquiring the evidence, especially when a survey questionnaire is used. Module 3 provides tips for writing valid and effective survey questions.

When you critically appraise the experience and/or professional judgment of a practitioner, you can use the checklist below to grade its trustworthiness. Keep in mind that this is not an exact science—determining whether or not there could have been bias is often a matter of interpretation and requires careful consideration of the context and circumstances.

Overview of appraisal questions

	Yes	No	Unclear
1. Is the domain or activity in which the practitioner gained experience sufficiently narrow and specific?			
2. Does the practitioner have extensive experience with the matter (problem/solution)?			
3. Was the practitioner able to evaluate the outcome, and if so, was direct, objective feedback available?			
4. Can the organizational context in which the practitioner gained his or her experience be regarded as sufficiently regular and predictable?			
5. If applicable, has an effort been made to reduce bias by taking measures? (Consider blind assessment, falsification of views and judgments seeking disagreement, introducing an opposite view, devil's advocate)			
6. To what extent could cognitive bias have affected the practitioner's judgment? (Consider social desirability bias, patternicity/illusion of causality, confirmation bias, group conformity, availability bias, authority bias, outcome bias, overconfidence bias			
7. Was the evidence acquired in a valid and reliable way?			

Did I Get This 4.11

Read the following scenario

The board of an organization that provides substitute teachers for primary schools has noticed that in recent years, fewer substitutes are available. Where they once had a choice among several substitutes, it now takes days, and sometimes even weeks, to find a suitable substitute teacher.

When they consult with two of the organization's substitutes, the substitutes suggest that the drop in availability may be due to the increasing number of children with severe emotional and behavioral problems in classrooms. Both workers recently completed their training and have been with the organization for about six months. They have not spoken directly with other substitutes about this issue but were surprised to find that the classes they stand in for include so many children with emotional and behavioral problems. Surely, they believe, this must affect potential substitutes' motivation to teach a class of children on a substitute basis.

How would you, based on this information, grade the trustworthiness of the substitutes' judgment?

a. Very high
b. -
c. -
d. -
e. Very low

Did I Get This 4.12

Read the following scenario

A medical student has just started their clinical internship in the urgent care department. During their shift, they encounter a 2-year-old toddler whose mother insists that something is wrong because the child is behaving differently than usual. However, all diagnostic tests come back negative. Despite this, the mother continues to believe that something is wrong with her child.

The medical student consults the senior pediatrician, who strongly advises admitting the child for observation. When the student asks for the rationale behind this advice, the pediatrician explains that mothers spend nearly 24 hours a day with their children and are therefore highly attuned to any changes in their behavior. In fact, based on decades of clinical experience, it has been observed that when a mother insists something is wrong with her child, even when diagnostic tests are negative, she is almost always right, and there is indeed a medical condition. For this reason, the international medical guideline recommends admitting the child for observation in such cases.

How would you, based on this information, grade the trustworthiness of the pediatrician's advice?

a. Very high
b. -
c. -
d. -
e. Very low

First we explained what professional expertise is: the specialized knowledge acquired by repeated experience and practice of specialized activities. However, accumulated experience alone does not necessarily result in reliable expertise—only under four specific circumstances does professional experience lead to valid and reliable expertise: 1) When the domain or activity is sufficiently narrow and specific; 2) When there are opportunities for repeated practice;3) When practice leads to direct, objective feedback; 4) Within a regular, predictable work environment. The less these four conditions are met, the more likely a professional's expertise is to be biased.

We also discussed why it is hard to develop valid and reliable professional expertise. Our professional judgment is highly susceptible to systematic errors – cognitive and information-processing limits make us prone to biases that have negative effects on the quality of the decisions we make. To illuminate how the brain works when making decisions, we explained the difference between System 1 and System 2 thinking.

- System 2 is slow, effortful, deliberate, and rational. It is the slow, effortful reasoning system that draws heavily on our cognitive resources and requires attention and concentration.
- In contrast, System 1 is fast, intuitive, associative, and emotional. It is the fast, effortless thinking system that operates automatically with little voluntary control and that uses intuition or heuristics to make decisions quickly. However, the systematic errors – also called *cognitive biases* – that affect how we make decisions stem directly from our System 1.

Among the many different types of cognitive biases are the following:

- **Patternicity/illusion of causality:** Seeing order, patterns, and causal relations in the world
- **Confirmation bias:** Confirming our existing beliefs by selectively sea ching for and interpreting information in a supporting fashion while ignoring information to the contrary
- **Group conformity:** Conforming to the others in a group, even if doing so goes against one's own judgment
- **Availability bias:** Relying on examples that spontaneously come to mind when evaluating a specifi topic, method, or decision
- **Authority bias:** Overvaluing the opinion of someone who is seen as an authority
- **Outcome bias:** Evaluating the quality of a decision on the basis of its outcome
- **Overconfidence bias:** Having an unwarranted faith in one's own knowledge, judgments, cognitive abilities, or skills

Unfortunately, being aware that human judgment suffers from cognitive biases does not prevent biases from occurring. In fact, as an individual, it is very hard to eliminate the biases in yourself. However, they can be neutralized – or at least reduced – at the organizational level by applying an evidence-based approach: checking the judgments that are made by acquiring, appraising, and applying multiple sources of evidence. In addition, practical measures can be taken to prevent cognitive biases from clouding the decision-making process:

- Collating the evidence before, not after, forming an opinion
- Performing blind assessments
- Falsifying views and judgments
- Seeking disagreement
- Playing devil's advocate
- Installing a red team

Podcast

In this podcast host Karen Plum discusses with Rob Briner, Professor of Organizational Psychology at Queen Mary University of London, Eric Barends, Managing Director of CEBMa, and Christina Rader, Professor at Colorado College, the quality and reliability of the evidence from practitioners – people who have experience with the problem we are tackling, or the solutions we are considering.

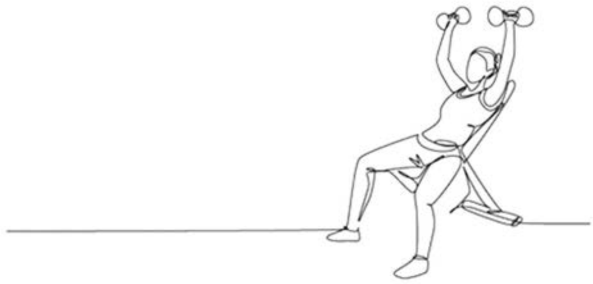
We consider the role of the evidence-based practitioner and the challenges of getting a clear understanding of problems and solutions from practitioners who may not always be as clear or succinct as we might wish. This is where careful listening and questioning are absolutely critical, together with challenging assumptions - both our own and other people's.

Finally, we discuss the three criteria used to assess the reliability and trustworthiness of practitioners and experts, including consultants that seek to identify themselves as experts with particular problems and solutions.

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 4.1: Cognitive biases

During a (team-, executive-, project-) meeting in your organization, take the role of observer and consider how people at the meeting are expressing their professional expertise and judgment. Reflect on the extent to which the information they provide may be biased.

1. Describe the cognitive bias that may have influenced the participants' judgment
2. How could this bias affect the decision-making process?
3. What could be done to avoid/mitigate this bias?

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

Exercise 4.2: Personal Reflection

Think of a recent professional decision you made. Reflect on the factors that influenced your judgment and consider whether any cognitive biases played a role.

1. Identify and describe any cognitive bias that may have influenced your decision.
2. Explain how this bias might have affected the outcome.
3. Consider what you could do differently in the future to reduce bias in similar situations.

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

Exercise 4.3: Organizational Bias

Rather than focusing on a single meeting, take a broader perspective and analyze your organization's decision-making culture. Observe patterns in discussions, communications, policy papers, and decisions made by managers, executives, or the board that suggest a dominant cognitive bias influencing the organization's way of thinking.

1. Identify and describe a cognitive bias that appears to be particularly prevalent or characteristic of your organization's decision-making processes. Provide specific examples or observations that support your analysis.
2. Explain how this bias may shape decisions, impact outcomes, or reinforce existing organizational norms.
3. Suggest practical strategies the organization could implement to recognize, counteract, or reduce the influence of this bias.

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Exercise 4.4: Strength of the evidence

During a (team-, executive-, project-) meeting in your organization, take the role of observer and monitor whether a judgment, opinion, claim, or recommendation is expressed by (one or more) participants. Respectfully ask the participant(s) what the evidence is for this judgment/ recommendation. If the source of evidence is 'professional experience/ expertise' determine whether the three criteria for valid and reliable expertise are met.

1. **Numerous opportunities to practice:** Does the person(s) have extensive experience with the matter (problem, preferred solution)?
2. **Direct, objective feedback:** Was the person(s) able to evaluate the outcome, and if so, was direct, objective feedback available?
3. **A regular, predictable work environment:** Can the organizational context in which the person(s) gained his or her experience with the matter be regarded as regular and predictable?
4. How would you, based on your answer to these questions, rate the trustworthiness of the person(s) judgment/recommendation?

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Suggestions for further reading

- *Judgment under uncertainty: Heuristics and biases*, Tversky & Kahneman, Science, 1974
- *Interview with Daniel Kahneman*, Strategy and Business, 2003
- *The Enron board: The perils of groupthink*, Marleen O'Connor, University of Cincinnati Law Review, 2003
- *The Delphi method for graduate research*, Skulmoski et al, JITE-Research, 2007
- *Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy*, Yaniv & Choshen, Journal of Behavioral Decision Making, 2012
- *Experience matters? The impact of prior CEO experience on firm performance*, Hamori & Koyuncu, HRM 2015

References

1. Dobelli, R. (2015). *The art of thinking clearly*. HarperCollins. New York, NY.
2. Ford, L. (October 3, 2005). “Australian pair wins Nobel prize for stomach ulcer research.” *The Guardian*.
3. Gladwell, M. (2002). “The talent myth: Are smart people overrated?.” *The New Yorker*. Volume 22. 28-33.
4. Tracey, T. J., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). “Expertise in psychotherapy: An elusive goal?.” *American Psychologist*. Volume 69. Number 3. 218.
5. Harvey, N. (2011). Learning judgment and decision making from feedback. In M. K. Dhami, A. Schlottmann, & M. R. Waldmann (Eds.), *Judgment and decision making as a skill: Learning, development, and evolution*. Cambridge University Press. Cambridge, UK. 199-226.
6. Kohler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biase beyond the laboratory. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment.* Cambridge University Press. New York, NY. 686-715.
7. Hamori, M., & Koyuncu, B. (2015). “Experience matters? The impact of prior CEO experience on fir performance.” *Human Resource Management*. Volume 54. Number 1. 23-44.
8. Spengler, P. M., & Pilipis, L. A. (2015). “A comprehensive meta-reanalysis of the robustness of the experience-accuracy effect in clinical judgment.” *Journal of Counseling Psychology*. Volume 62. Number 3. 360-378.
9. Tversky, A., & Kahneman, D. (1974). “Judgment under uncertainty: Heuristics and biases.” *Science*. Volume 185. Number 4157. 1124-1131.
10. Kahneman, D. (2011). *Thinking, fast and slow.* Penguin Group. London.
11. Schrage, M. (2003). “David Kahneman: The thought leader interview.” *Strategy+Business*. Volume 33.
12. Adapted from Gilovich, T. (2008). *How we know what isn’t so*. Simon and Schuster.
13. Shermer, M. (2011). *The believing brain: From ghosts and gods to politics and conspiracies*. Time Books. New York, NY.
14. Krebs, V. (2008). *Political polarization during the 2008 US presidential campaign.* Orgnet. http://www.orgnet.com/divided.html.
15. Lauren(November 28, 2000). *Blind orchestra auditions better for women, study finds.* Princeton University All News. https://www.princeton.edu/news/2000/11/28/blind-orchestra-auditions-better-women-study-finds.
16. Goldin, C., & Rouse, C. (2000). “Orchestrating impartiality: The impact of “blind” auditions on female musicians.” *American Economic Review*. Volume 90. Number 4. 715-741.
17. Maddux, W. W., Swaab, R. I., Tanure, B., & Williams, E. (2014). “*Ricardo Semler: A revolutionary model of leadership.*” INSEAD Case #04/2014-5982.
18. McRaney, D. (2011). *You are not so smart*. Gotham Books?Penguin Group. New York, NY.
19. Pande, P. S., Neuman, R. P., & Cavanagh, R. R. (2000). *The Six Sigma way*. McGraw-Hill. New York, NY.
20. Pepper, D.(May 9, 2016). *How 3M lost (and found) its innovation mojo* https://www.inc.com/linkedin/don-peppers/downside-six-sigma-don-peppers.html.
21. Morris, B.(July 10, 2006). “New rule: Look out, not in.” *Fortune* [Archive]. https://archive.fortune.com/2006/07/10/magazines/fortune/rule4.fortune/index.htm.
22. Larsen, K. S. (1990). “The Asch conformity experiment: Replication and transhistorical comparisons.” *Journal of Social Behavior & Personality*. Volume 5. Number 4. 163-168.

23. Hermann, A., & Rammal, H. G. (2007). “The grounding of the ‘flying bank’.” *Management Decision*. Volume 48. Number 7. 1048-4062.
24. O’Connor, M. A. (2003). “The Enron board: The perils of groupthink.” *University of Cincinnati Law Review*. Volume 71. Number 4. 1223-1320.
25. Abrahamson, E. (1991). “Managerial fads and fashions: The diffusion and rejection of innovation.” *Academy of Management Review*. Volume 16. Number 3. 586-612.
26. Staw, B. M., & Epstein, L. D. (2000). “What bandwagons bring: Effects of popular management techniques on corporate performance, reputation, and CEO pay.” *Administrative Science Quarterly*. Volume 45. Number 3. 523-556.
27. Colven, G. (June 28, 2004). “A concise history of management hooey: For every Six Sigma quality initiative, there’s an inkblot test.” *Fortune* [Archive]. https://archive.fortune.com/magazines/fortune/fortune_archive/2004/06/28/374405/index.htm.
28. Darwin, C. (1859). *On the origin of species*. Gryphon Editions. Birmingham, AL.
29. Parks, W.-W. (1990). “A review of research on groupthink.” *Journal of Behavioral Decision Making*. Volume 3. Number 4. 229-245.
30. Esser, J. K. (1998). “Alive and well after 25 years: A review of groupthink research.” *Organizational Behavior and Human Decision Processes*. Volume 73. Number 2. 116-141.
31. Kaushik, A. (2007). *Web analytics: An hour a day*. Wiley. Indianapolis, IN.
32. Svenson, O. (1981). “Are we all less risky and more skillful than our fellow drivers?.” *Acta Psychologica*. Volume 47. Number 2. 143–148.
33. Zuckerman, E. W., & Jost, J. T. (April 24, 2000). “*It’s academic.*” Stanford GSB Reporter. 14-15.
34. Lambert, M. J. (2010). Prevention of treatment failure: *The use of measuring, monitoring, and feedback in clinical practice*. American Psychological Association. Washington, DC.
35. Kruger, J., & Dunning, D. (1999). “Unskilled and unaware of it: How difficulties in ecognizing one’s own incompetence lead to inflated self-assessments.” *Journal of Personality and Social Psychology*. Volume 77. Number 6. 1121-1134.
36. Montier, J. (2007). *Behaving badly. In Behavioural investing: A practitioner’s guide to applying behavioural finance*. Chapter 7. Wiley. Hoboken, NJ.
37. Kahneman, D., Lovallo, D., & Sibony, O. (2011). “Before you make that big decision.” *Harvard Business Review*. Volume 89. Number 6. 50-60.
38. Doyle, A. C., & Holt, R. (1999). *A scandal in Bohemia*. Pearson Education. Harlow, UK.
39. Sloan, A. P. (1964). *My years with general motors*. Doubleday. Garden City, NY.
40. Herbert, T. T., & Estes, R. W. (1977). “Improving executive decisions by formalizing dissent: The corporate devil’s advocate.” *Academy of Management Review*. Volume 2. Number 4. 662-667.
41. Kahneman, D. (2003). “A perspective on judgment and choice: Mapping bounded rationality.” *American Psychologist*. Volume 58. Number 9. 697-720.
42. Nisbett, R. E. (Ed.) (1993). *Rules for reasoning*. Erlbaum. Hillsdale, NJ.
43. Larrick, R. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making*. Blackwell. Malden, MA.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 4.1

Answer is provided in the module.

Learn by doing 4.2

A good example is a sales manager. In general, selling a product or service involves a specific activity. Sales managers work within a relatively steady and predictable work environment, they give their sales pitch several times a week (or even day), and they receive frequent, direct, and objective feedback: the deal is accepted or is not accepted.

Did I get this? 4.1

A carpenter specializing in wooden stairs uses fairly standardized methods—the sizes and placement of the stairs will vary, but the way the stairs are measured, designed, and crafted will be fairly standard. When the carpenter makes an error (e.g., a measurement error), he or she will directly notice the consequences (the stairs don't fit). In addition, a carpenter specializing in wooden stairs will craft stairs every week, allowing their professional expertise to develop through prolonged practice of a specific activity in a consistent and predictable environment. For this reason, we would rank the carpenter's professional expertise as the most reliable.

A clinical psychologist who is specialized in anxiety disorders works in a regular and predictable environment, and although the treatment of patients is individualized, the psychologist can often use a standardized treatment protocol based on evidence-based guidelines. In addition, the psychologist will see several patients per day, and the outcome of the treatment can be (subjectively) judged not only by the psychologist and patient but also (objectively) through use of a validated rating system measuring the severity of anxiety symptoms. For this reason, we would rank the psychologist's expertise as second.

A trader at the stock exchange who is specialized in gold stocks makes several trades a day, and the feedback he or she receives is fairly direct and objective (the stock goes up or down). However, the trader works in a very dynamic and unpredictable environment. Gold stock prices are wildly volatile—they are influenced not only by the price of gold but also by political events and macroeconomic factors. For this reason, we would rank the trader's expertise rather low.

Finally, the work of an HR manager specialized in people management encompasses a rather broad domain and wide range activities (e.g. recruitment and selection, talent management, coaching, team building, learning and development, etc.). Additionally, they work in the same dynamic environment as the management consultant mentioned earlier. Finally, the HR manager receives limited feedback, partly because the failure or success of people management is often hard to measure. For this reason, we would rank the HR manager's expertise as the least reliable.

Learn by doing 4.3

Our answer: Most people are inclined to answer "10 cents" because the sum $1.10 so neatly separates into 1 dollar and 10 cents. In fact, more than half of a group of students at Princeton and at the University of Michigan gave precisely that answer – the wrong answer. The correct answer is that the ball costs 5 cents. However, it takes some computational effort to come up with the right answer. As a result, our "lazy" System 2 is inclined to take the back seat and allow System 1 to quickly come up with an answer that sounds superficially plausible but is actually w ong.

Learn by doing 4.4

1: System 1 > It appears that the senior managers have fallen prey to the halo effect heuristic: the presentation looks professional, and the proposed model is taken at face value rather than critically evaluated. This suggests that the decision to hire the consultant and adopt the model was based on System 1 rather than System 2 thinking.

2: System 2 > The administrator could have fallen prey to the wisdom of the crowd heuristic: if most physicians think the new system is a good idea, it will probably work. It is unknown, however, whether the physicians have any experience with the system or that their judgment is based on only an assumption. His decision to find out what the experiences a e of other hospitals is therefore an example of System 2 thinking.

3: System 2 > The fact that the recommendation was made by a well-known, prestigious international consulting firm appears not to have a fected the CEO's decision, as she decided to consult the scientific literatu e first. This suggests that her decision not to int oduce performance bonuses (yet) was based on System 2 rather than System 1 thinking.

4: System 2 > By anonymizing the proposals, the CEO ensures that the people who will assess the proposals will not make a decision based the familiarity heuristic: choosing the consulting firm whose name they are familiar with.

Did I get this? 4.2

System 1 > We would argue that the CEO's decision to adopt agile working is not based on a critical evaluation of this new method but rather on the wisdom of the crowd heuristic: if so many say/think it's good, it must be good. This suggests that his decision was based on System 1 rather than System 2 thinking.

Learn by doing 4.5

1. The answer is provided in the module.
2. According to many people the bun resembles Mother Theresa.

Learn by doing 4.6

The answer is provided in the module.

Did I get this? 4.3

The answer is provided in the module.

Did I get this? 4.4

1: Incorrect > A video of the applicant won't help prevent confirmation bias, because all information that could induce bias is still present.

2: Incorrect > Recruiters from an external firm cannot be egarded as less prone to bias than people in the selection committee of a company.

3: Correct > *Blinding* is a widely used method to prevent confirmation bias. Hiding all information in a person's curriculum vitae and application letter that could induce bias, such as age, gender, and college attended, allows people in the selection committee to make an objective assessment of a candidate's qualities.

4: Not quite right > Leaving out an applicant's photo will surely help, but unfortunately, it is not enough, as information in the application letter that could induce bias is still present.

Did I get this? 4.5

1: Incorrect > The documents that will be sent by Semler's company will obviously paint a very positive picture, which will confirm the CEO in his positive view of the model

2: Incorrect > The CEO may ask critical questions during his visit, but it is unlikely that he will get objective, unbiased answers.

3: Correct > Actively looking for evidence that contradicts your views and judgments helps to overcome confirmation bias

4: Incorrect > Aside from the question whether such evidence would be available, this evidence would only confirm your CEO s positive view of the model rather than question it.

Did I get this? 4.6

Our answer: Some empirical studies suggest that an overly high level of demographic homogeneity within a board may cause its members to be especially prone to conformity bias. And although group homogeneity is not the only antecedent of group conformity, the quality of decision-making would likely benefit f om a more heterogeneous composition in terms of age, race, gender, and social background.

Learn by doing 4.7

1a: Incorrect > Group conformity is the tendency to conform to the others in a group, even if doing so goes against your own judgment. In this example, there is no group to conform to.

1b: Incorrect > Authority bias is the tendency to overvalue the opinion of someone who is seen as an authority. In this example, there is no person whose status and credentials influence to decision to choose surgeon B.

1c: Incorrect > Confirmation bias is the tendency to confirm our existing beliefs by selectively se ching for and interpreting information in a supporting fashion while ignoring information to the contrary. You could argue that people have an existing belief that the doctor with the best medical outcome is the best doctor and therefore choose surgeon B, but that is a different bias.

1d: Correct > Outcome bias is the tendency to evaluate the quality of a decision (intervention, method) on the basis of its outcome, in this case a successful operation.

1e: Incorrect > Availability bias is the tendency to rely on examples that spontaneously come to mind when evaluating a specific decision. In this case, the decision to choose surgeon B is not influence by examples that spontaneously come to mind.

1f: Incorrect > Patternicity/illusion of causality is the tendency to see order, patterns, and causal relations. In this example, there is no perceived pattern or assumed causal relation.

1g: Incorrect > Overconfidence bias is the tendency to have an unwarranted faith in one s own knowledge, judgments, cognitive abilities, or skills. In this example, it is not about one's own knowledge and skills but about those of other people (surgeon A and surgeon B).

2: The sample (n = 1) is too small, and you can judge a surgeon's professional expertise only if you know something about the circumstances, the patient's condition, the surgeon's level of experience, and the way the operation was executed.

3: How experienced is the surgeon? How many times has he/she performed this operation? What was the outcome of his/her operations over the past year? What was the condition of the patient during the last operation?

Learn by doing 4.8

1a: Yes > Your friend seems to have discovered a pattern: you can predict the price of gold by multiplying the position of the Dow Jones index by the oil price. Human beings, however, are predisposed to see meaningful patterns in both meaningful and meaningless noise. He therefore may have fallen prey to patternicity.

1b: No > Confirmation bias is the tendency to confirm our existing beliefs by selectively se ching for and interpreting information in a supporting fashion while ignoring information to the contrary. While you could argue that your friend has an existing belief (the gold price and the Down Jones index are closely correlated) and ignores information to the contrary, but this is triggered first and fo emost because he (thinks he) sees a conspicuous pattern.

1c: No > Group conformity is the tendency to conform to the others in a group, even if doing so goes against your own judgment. In this example, there is no group to conform to.

1d: No > Availability is the tendency to rely on examples that spontaneously come to mind when evaluating a specific decision. our friend's belief that the gold price and the Dow Jones index are closely correlated is not influenced by examples that spontaneously come to mind

1e: No > Authority bias is the tendency to over-value the opinion of someone who is seen as an authority. In this example, there is no person whose status and credentials influence your friend s belief that the gold price and the Dow Jones index are closely correlated.

1f: No > Outcome bias is the tendency to evaluate the quality of a decision (intervention, method) on the basis of its outcome. In this example, there is no specific outcome that influences your frien s judgment.

1g: No > Overconfidence bias is the tendency to have an unwarranted faith in one s own knowledge, judgments, cognitive abilities, or skills. It could be argued your friend has great faith in his own judgment, but this is triggered by the fact that he truly believes the gold price and the Dow Jones index are closely correlated.

2. We would ask, In the past six months, how often did this phenomenon occur? In addition, we would suggest monitoring the gold price for the next three months to see how robust the phenomenon is.

Did I get this? 4.7

a: No > Patternicity/illusion of causality is the tendency to see order, patterns, and causal relations. In this example, there is no perceived pattern or assumed causal relation.

b: No > Confirmation bias is the tendency to confirm our existing beliefs by selectively se ching for and interpreting information in a supporting fashion while ignoring information to the contrary. In this example, the managers did not selectively search for information in support of their existing beliefs.

c: No > Group conformity is the tendency to conform to the others in a group, even if doing so goes against your own judgment. In this example, there is no group to conform to.

d: No > Availability bias is the tendency to rely on examples that spontaneously come to mind when evaluating a specific decision e: No > Authority bias is the tendency to over-value the opinion of someone who is seen as an authority. In this example, there is no person whose status and credentials influence the managers' confidenc in their own performance.

f: No > Outcome bias is the tendency to evaluate the quality of a decision (intervention, method) on the basis of its outcome. In this example, there is no specific outcome that influences the managers judgment.

g: Yes > The majority of managers seem to have an unwarranted faith in their own performance.

Did I get this? 4.8

a: Yes > You could argue that people saw a pattern: every year, the housing prices increase by about 10 percent.

b: Yes > It seems their judgment could have been affected by confirmation bias. They st ongly believed that the housing prices would continue to increase, and as a result, they assigned more weight to evidence that confirmed this belief and igno ed the signs that the housing market was slowing down.

c: No > You could argue that many people believed that prices would continue to increase by 10 percent a year because most other people believed so too, but this is more a case of relying on the wisdom-of-the-crowd heuristic and is less about conforming to the group.

d: No > In this example, people's belief that prices would continue to increase by 10 percent a year was not influenced by examples that spontaneously came to mind

e: No > However, it could be argued that people's belief that prices would continue to increase by 10 percent a year was influenced by authoritative people such as Federal Reserve Chairman Ben Bernanke.

f: No > In this example, there seems to be no specific outcome that influenced peoples' judgmen

g: Yes > The financial experts may have been p one to overconfidence bias, the tendency to have an unwarranted faith in one's own knowledge or judgment.

Learn by doing 4.9

a: No > A person playing devil's advocate and taking an opposite view could determine the validity of your CEO's enthusiasm about the new IT system and identify biases in her argumentation, thereby increasing the quality of the decision-making process. However, it would be hard for this person to play devil's advocate during a visit at a company that has just successfully implemented the new system, which is why this is not the best advice.

b: No > While installing a red team to challenge your CEO's assumptions and to identify biases might be helpful, red teams are often installed to challenge (project) teams, rather than individuals, which is why it is not the best advice.

c: Yes > Our System 1 strongly favors our own ideas, so advising your CEO to look for evidence that may contradict her view may help her overcome her biases. In this example, the best advice you could give is to seek out and visiting a company where the implementation of this new IT system was *not* successful.

d: No > Blinding is an effective method to prevent confirmation bias f om occurring when someone holds a strong preference for one option over another. In this case, the CEO considers only one option – a specific IT system – so blinding would not help

Did I get this? 4.9

1a: Incorrect > While you might argue that the senior executives suffer from overconfidence bias because they are confident that it is lack of risk awa eness that is causing the errors, that is not the bias they would be prone to by discussing their view with the middle managers.

1b: Incorrect > While you might argue that a few of the senior executives believed they knew what was causing the errors and their opinions may have influenced the other senior executives, that is not the bias they would be prone to by discussing their view with the middle managers.

1c: Correct > Discussing their view with the middle managers to see whether they agree makes them prone to confirmation bias

2a: Yes > Installing a red team to challenge the executives' opinion and identify biases will help the executives make a better decision.

2b: Yes > By assigning someone as devil's advocate, you could determine the validity of the executives' judgment that the errors are due to a lack of risk awareness.

2c: The executives have decided to discuss their view with some middle managers to see whether they agree, which makes the executives prone to confirmation bias. Instead, they should consult people within the organization who may have contradictory beliefs.

3: The senior managers should actively seek information that contradicts their assumption that errors are being made because of employees' lack of risk awareness. They also should consult people within the organization who may have contradictory beliefs. In addition, they could consult the scientific literatu e to see whether lack of risk awareness is indeed a common cause of people making mistakes or whether other causes are more likely.

Did I get this? 4.10

Our answer: Elements that you should consider are the composition of the team (age, race, gender, social background, area of expertise) and a good decision-making process (extensive search and appraisal of information, a thorough consideration of alternatives). In addition, the team should be encouraged to disagree, for instance, by appointing a devil's advocate, and should collate all the appropriate evidence before forming an opinion, not after. Finally, you could consider setting up a second team working on the same problem.

Did I get this? 4.11

Both substitute teachers have very limited experience, both in the profession (having just graduated) and within the organization (only 6 months). Additionally, they are unable to test their hypothesis—that nowadays more children have severe emotional and behavioral problems—in a direct, objective manner. Due to their limited work experience, they cannot compare the number of children with emotional and behavioral problems in classrooms now to a few years ago. Furthermore, they have not spoken with other substitutes to verify whether their experiences are common. Finally, we cannot rule out the possibility that bias has influenced their judgment: their experience is based on only a few classes (selection bias), and they may have held the (false) expectation that children in class are always friendly and cooperative. Based on this, we would grade the trustworthiness of their judgment as low to very low.

Did I get this? 4.12

Senior pediatricians have many years of clinical experience and are often able to evaluate the outcomes of their decisions in a direct, objective way. In this case, either the diagnostic tests were wrong and the mother was right, or there was indeed nothing wrong with the child. Additionally, it is unlikely that the organizational context and dynamics have affected their judgment. Finally, the pediatrician's advice aligns with the clinical guideline—clinical guidelines are typically based on a large evidence base, so it is unlikely that their advice is affected by bias.

.

Based on this, we would grade the trustworthiness of the pediatrician’s judgment as very high.

.

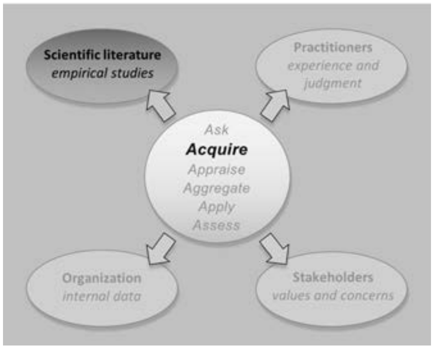
MODULE 5 | ACQUIRE Evidence from the Scientific Literature

Learning objectives:

- Determine the best place to search for empirical studies relevant to the question.
- Identify effective search terms to search for relevant empirical studies in an online research database.
- Conduct a systematic and reproducible search for empirical studies in online research databases.

Introduction

> *Our duty is to believe that for which we have sufficient evidence and to suspend our judgment when we have not.*
>
> *– JOHN LUBBOCK*



Consider the following real-life example.

A consulting company developed a new model to ensure better accountability, conduct, and culture in banking firms. When asked about the scientific evidence underlying the model, the company provided the following answer:

Our model is grounded in the work of Professor O'Neil on trust and trustworthiness. In addition, we explored the scientific literature, a subset of which includes the seminal work of Professor Schein on the role of culture in financial services, Professor Coates and others on biological drivers, Professor Smith from Harvard Business School on culture and dishonesty in banking, Professor Rousseau from Carnegie Mellon University on measuring and assessing culture in healthcare, Professor Edmondson on psychological safety, and Professor Johnson from the London Business School on moral disengagement and unethical behavior.

We hope that this clarifies the evidence-based approach we have taken, whereby in aiming to build a cutting-edge model we are informed by the scientific literature and guided by leading academics from several renowned universities and business schools.

Would you consider this answer to be a good example of an evidence-based approach, in particular, consulting the evidence from the scientific literature? After all, the firm consulted several academics and asked for their views on the evidence and probably looked at some of the scientific literature too.

Consulting some well-known professors from well-known universities and looking at just some of the scientific literature, however, is not technically an evidence-based approach. The entire purpose of evidence-based management is to get away both from consulting experts about evidence and their opinion of such (which is regarded as low-quality evidence in evidence-based management) and from dipping into the scientific literature. We should consult the scientific literature in a systematic, transparent, and verifiable way; our aim is to minimize bias by having an explicit search strategy and clear criteria for judging the methodological quality of the evidence we find. In this sense, scientific evidence has nothing to do with the opinions of experts or the work of particular researchers – our search is for all the relevant scientific evidence, which we then judge using objective criteria

So, yes, we agree that the consulting firm consulted the scientific literature, but this is not the same as taking an evidence-based approach. To bring evidence from the scientific literature into your decisions, you need to know how to search for empirical studies in online research databases. This module therefore teaches you the skills necessary to successfully conduct a systematic, transparent, and verifiable search using online research databases such as ABI/INFORM Global, Business Source Premier, and PsycINFO.

In this module we provide a link to the database Business Source Elite.

Before you click this link, **please make sure that you don't use an internet connection (e.g. WIFI or intranet) through a university or any other educational institution**, as this may result in different search outcomes!

Where Do We Search?

Academic Journals vs Popular Magazines

By evidence from the scientific literature, we mean empirical studies published in academic or scholarly journals. Especially in the field of management and business, popular magazines often make claims about how to lead an organization, grow a business, or manage people. However, popular magazines are very different from academic journals and are generally considered unreliable sources of evidence. The main differences are:

Academic (Scholarly) Journals	Popular Magazines
Articles are written by academics, scholars, or researchers.	Articles are written by business leaders, famous consultants, or so-called 'experts'.
Articles present the results of empirical research and/or scientific theory.	Articles present new insights, trends, best practices, and/or the opinions of the author(s).
Articles are peer-reviewed by other academics or scholars within the field.	Articles are reviewed by an editor of the magazine.
The vocabulary is technical, academic and often difficult to read.	Common and entertaining language is often used, making the articles easy to read.
Graphics are minimal and tables are technical and specialized.	There are multiple images and illustrations.
Articles include citations and a bibliography referencing the sources used.	There are no citations and often limited (or no) information about the sources used.

The Process of Peer Review

Articles submitted to academic journals are first evaluated and critiqued by independent, anonymous scientists in the same field (peers) to determine whether they merit publication in a scientific journal. Under this approach, an author can revise the article to make corrections and include any peer reviewers' suggestions that will make the article stronger, such as incorporating previously overlooked ideas and addressing methodological concerns. If the author cannot or will not take the peer reviewers' advice, the article may be rejected for publication.

Peer review ensures that an article—and therefore the journal and the discipline as a whole—maintains a high standard of quality, accuracy, and academic integrity. Of course, this sounds good in theory, but in practice, it is not always the case: poor-quality studies suffering from methodological flaws, bias, and incomplete conclusions are sometimes rife in peer-reviewed academic journals, even in top journals with a high impact factor (a measure reflecting that the articles in the journal are often cited by other researchers). For this reason, an evidence-based approach always involves critically appraising the studies found, even those published in well-known, highly reputed, peer-reviewed journals.

Did I get this 5.1

Read the following four statements regarding peer-reviewed journals. Determine whether the statement is true or false.

1. Research articles published in peer-reviewed journals are evaluated and critiqued by independent, anonymous scientists in the same field
2. The process of peer review ensures that a research article is valid and reliable.
3. A research article published in a peer-reviewed journal with a high impact factor does not need to be critically appraised.
4. Peer review does not give any guarantee that a research article is valid and reliable.

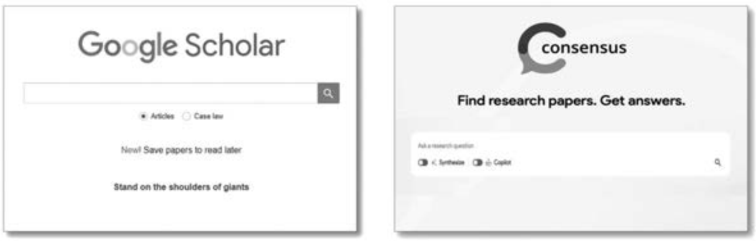
The easiest way to find peer-reviewed articles is to search in an online research database. To make sure your results come from peer-reviewed journals, you simply check the box that allows you to limit your results to peer-reviewed only.

Online research databases

As you have just learned, when we search for empirical studies, we first look for studies in peer-reviewed journals. In the past, this task entailed asking an academic or a business librarian for titles of journals likely to publish studies relevant to your question, then going to the library and sifting through tens to hundreds of issues until you found a sufficient number of studies. Today, a visit to the library is no longer necessary because most published research is retrievable through the Internet. In addition, online research databases make it possible to simultaneously conduct a search in thousands of peer-reviewed journals. In fact, you can now conduct a search for studies from any place at any time, for example, from home, from your favorite coffee shop, or during the break of an important business meeting.

There are literally hundreds of online research databases. Medicine, social welfare, biology, economy, and physics—every discipline has its own specialized research database. This means that if you want to answer a question that is relevant to the field of management, you should conduct your search in databases relevant to the field of management, such as ABI/INFORM or Business Source Elite. Depending on your question, you may also need to search in databases that are aimed at neighboring disciplines, such as psychology (PsycINFO), education (ERIC), or healthcare (PubMed)

Starting Point: Google Scholar and AI Search Tools



As explained, when we search for scientific evidence, we search in online research databases. However, as we will discus later in this module, this is not how we start. We start with a quick search in Google Scholar or with an AI search tool specialized in scientific research such as Consensus.

Google Scholar (https://scholar.google.com) is a web search engine that allows users to search for academic papers and scholarly articles across many disciplines. Google Scholar is easy to use because it has a similar design to the standard Google search engine. It is free to use and capable of finding academic articles that are not behind a paywall. Another benefit of Google Scholar is its intuitive search capability. For example, if you search for articles on "information sharing" in teams, it will suggest related articles. Google Scholar returns the most relevant results first, based on the articles' full text, author, source, and the number of times they have been cited in other articles.

AI search tools like Consensus go a step further than Google Scholar. If you search for "information sharing" in teams, it will also present studies on "knowledge sharing" in teams, as this is a related term. Additionally, Consensus allows you to filter your results by study design (e.g., meta-analyses or controlled studies) and automatically summarizes the results of the first five articles.

However, as we will discuss later in this module, Google Scholar and AI search tools have serious limitations. For this reason, an additional search in one or more research databases is often required.

What Do We Search For?

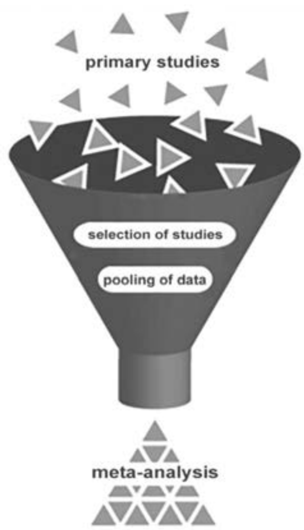
1. Empirical Research

As explained in the previous section, when we refer to evidence from the scientific literature, we mean empirical studies published in peer reviewed academic or scholarly journals

Empirical research involves gathering and analyzing real-world quantitative or qualitative data to gain knowledge about a phenomenon or intervention. Unlike studies that develop theories or concepts, empirical studies are based on real-life observations and measurements. For example, in management, instead of theorizing about how flexible work hours may increase employee productivity, an empirical study would involve observing companies that implement flexible hours, measuring employee productivity levels, and analyzing this data to draw conclusions about whether flexible hours indeed increase productivity. A special type of empirical research is a meta-analysis. In fact, when we search for empirical studies published in scholarly journals, we first look for meta-analyses.

2. Meta-analyses

When we search for empirical studies, we first search for meta-analyses. A meta-analysis is a study that statistically summarizes the results of a large number of empirical studies on the same topic. These empirical studies are referred to as `primary' studies because they are original investigations into a topic. You will learn more about meta-analyses in modules 6 and 7, but for now, it is important to know that a meta-analysis is essentially "a study of studies." The authors first conducted a rigorous search of primary studies on the same topic or research question (often in multiple online research databases), then selected the most relevant studies and critically evaluated them based on explicit criteria. They then statistically combined ('pooled') all the individual results into one overall outcome and summarized the findings. A similar type of study is a 'systematic review'. Although not technically the same, both summarize the results of multiple primary studies.



The reason to search for meta-analyses first is twofold:

1. The availability of meta-analyses indicates that the topic is well-researched, as the presence of multiple studies prompted researchers to conduct a meta-analysis, suggesting sufficient scientific evidence on the subject.

2. The authors of a meta-analysis or a systematic review have essentially done the work for you. They have searched multiple research databases, selected the most relevant studies, combined the results, and summarized the findings.

Meta-analyses and systematic reviews often provide the best available scientific evidence on a topic. Thus, if we find a systematic review or meta-analysis, we can save ourselves a lot of time and effort.

Section 1: Determining Your Search Terms

1.1. PICOC Terms

Before you start your search, it is important to make explicit the professional context that should be taken into account when answering your question. For example, questions such as “Does team-building work?” or “Does 360-degree feedback increase performance?” may be relevant, but they are also broad. You may be specifically interested in whether team-building improves product quality in a German manufacturing company that recently has undergone restructuring or whether 360-degree feedback is effective as a tool for improving governmental managers’ service to the public.

To make your question more context-specific, you can formulate a PICOC (Population, Intervention, Comparison, Outcome, Context). We discussed this tool in Module 2, ASK Critical Questions about Assumed Problems and Preferred Solutions. PICOC is a mnemonic to help you find studies that are relevant to your professional context.

Population	*Who?*	Type of employees
Intervention	*What or how?*	Management technique/method, factor, or independent variable
Comparison	*Compared to what?*	Alternative intervention, factor, or variable
Outcome	*What are you trying to accomplish, improve, or change?*	Objective, purpose, goal, or dependent variable
Context	*In what kind of organization or circumstances?*	Type of organization, sector, relevant contextual factors

Example

P: Physicians

I: 360-degree feedback

C: Coaching

O: Increased task performance

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

Example

P: Software developers

I: Agile working

C: Business as usual/status quo

O: Reduced software development costs

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

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 to answer the question 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 a e generalizable to all human beings, but sometimes what works in one narrowly defined setting may not work in another.

Learn by doing 5.1

Read the following scenario and determine the PICOC terms:

In the past four years, a large midwestern US manufacturing organization has experienced multiple restructurings and downsizings, consequently reducing its workforce from more than 8,000 to fewer than 7,000 factory workers. According to the HR director, the restructurings and downsizings have been very stressful for the workers and have led to fear of job loss and anxiety. He therefore recommends deploying a stress-reduction program that includes on-site chair massage therapy, a technique that has successfully been tried by several companies, including AT&T, Apple, and Google. Before the CEO decides whether to implement the program, however, he wants to know what is known in the scientific literature about the effect of chair massage therapy on the stress and anxiety levels of employees.

P: What is the population of interest?

a. Employees
b. Staff
c. Large US manufacturing organizations
d. Factory workers

I: What is the intervention (management technique, method) of interest?

a. Restructuring and downsizing
b. A stress-reduction program
c. Chair massage therapy
d. On-site chair massage therapy

C: What is the comparison (alternative method)?

a. A stress-reduction program
b. Restructuring and downsizing
c. Doing nothing/business as usual
d. Doing something to relax

O: What is the outcome of interest?

a. Absenteeism
b. Reduced stress and anxiety
c. Improved performance
d. Increased job satisfaction

C: What is the context (what kind of organization or circumstances)?

a. Any organization
b. A manufacturing organization
c. An organization that has experienced multiple restructurings and downsizings
d. A large midwestern US manufacturing organization that has experienced multiple restructurings and downsizings

Did I get this 5.2

Read the following scenario and determine the PICOC terms:

Hoping to imitate the innovative and flexible work environments found at startups and companies such as Google, the CEO of your organization, a small Italian software company, is considering implementing a bullpen style, open-plan layout. Currently, the office is divided into individual workspaces with half walls. With 15 software developers working in a relatively small space, you worry that the distractions created by a new, open layout may undermine workers' 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 your of office, you suggest that your CEO consult the scientific literature to find out what is known about the effect of open office designs on the concentration of employees.

Population

a. Software developers
b. Knowledge workers
c. Employees
d. Factory workers

Intervention

a. Flexible work environment
b. Office redesign
c. Innovative work environment
d. Open office design

Comparison

a. Doing nothing/status quo
b. Bullpen style office
c. Traditional office
d. Individual workspaces with half walls

Outcome

a. Distraction/concentration
b. Improved performance
c. Productivity
d. Job satisfaction

Context

a. A software company
b. A small Italian software company
c. An organization with a traditional office design
d. Any organization

1.2 The two most important PICOC terms

Determining your search terms is the most important step when searching for empirical studies in an online research database. As you can imagine, here too, the rule “garbage in, garbage out” (GiGo) applies. This is particularly important when you search in online research databases. Unlike Google Scholar and Consensus, research databases just process the search terms that are given. Good, clear, and specific search terms result in relevant outputs, whereas unclear, vague, ambiguous, or incorrect search terms will most certainly lead to irrelevant results.The following two steps will help you to identify search terms that will yield studies relevant to your question.

Step 1. Determine the two most important terms of your PICOC

The first step in finding relevant empirical studies is to determine what the two most relevant PICOC terms are. In most cases, this will be the intervention (management technique, independent variable) and the outcome (objective, outcome measure, dependent variable). Other PICOC terms, such as population and context, may also be important, but their specificity often leads to exclusion of relevant studies, so as a rule, we leave them out.

Learn by doing 5.2

Read the following scenario and determine the two most important PICOC terms.
Select Yes or No.

One of my workplace requirements is that each department must have a certified coach who will be able to provide coaching sessions when requested by other peers. Coaching has become quite ubiquitous, as the belief is that the clinical managers of our organization, an academic hospital, will enhance their performance if they endure coaching sessions. I would like to understand if scientific evidence supports this claim.

P: Clinical manager

P: Business coaches

I: Coaching

C: Academic hospital

O: Individual performance

Learn by doing 5.3

Read the following scenario and determine the two most important PICOC terms.

In the past three decades, the number of workers in healthcare engaged in knowledge work has substantially increased. Whether nurses, physicians, managers, or staff members, most workers in healthcare organizations today rely heavily on information and communication technology and are involved in work that is at a high level of cognitive activity. As a healthcare manager, I have a responsibility to optimize work processes and enhance my employees' performance. My question, therefore, is: What is known in the scientific literature about factors or variables that impact the performance of knowledge workers?

P: Knowledge workers

I: Factors or variables

O: Cognitive activity

O: Performance

C: Healthcare

Did I get this 5.3

Read the following scenario and determine which of the following are the two most important PICOC terms.

As a knowledge consultant, I was commissioned by a Dutch company that specialized in innovative furniture design to make the organization's design teams more innovative through enhanced knowledge sharing. I'm not sure, however, whether knowledge sharing has an effect on innovation, so I would like to know what is known in the scientific literature about this topic.

P: Design teams

I: Knowledge sharing

C: Lack of knowledge sharing

O: Team innovativeness

C: Dutch furniture design company

1.3. Management Jargon vs Academic Terms

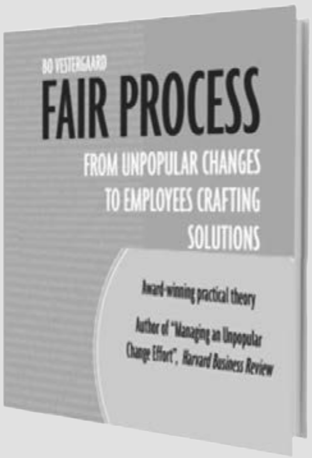
Step 2: Identifying corresponding academic terms

People often state that managers and academics live in very separate worlds. This is particularly true for the terms and jargon they use. For instance, for managers, performance is often just performance, whereas academics distinguish between many different types of performance, such as task performance, contextual performance, counterproductive work behavior, extra role performance, and organizational citizenship behavior. The same is true for terms that managers use, which sometimes have corresponding but different names in academia. As a result, searching for studies with managerial (i.e., non-academic) terms is unlikely to yield relevant results.

Example

The outcomes of an organizational change intervention can be both positive and negative, depending on the type of change and the specific individual or group affected. Especially when the change has predominantly negative outcomes (e.g., lay-offs), it is assumed to be important that the change process is perceived by the employees to be a 'fair process'.

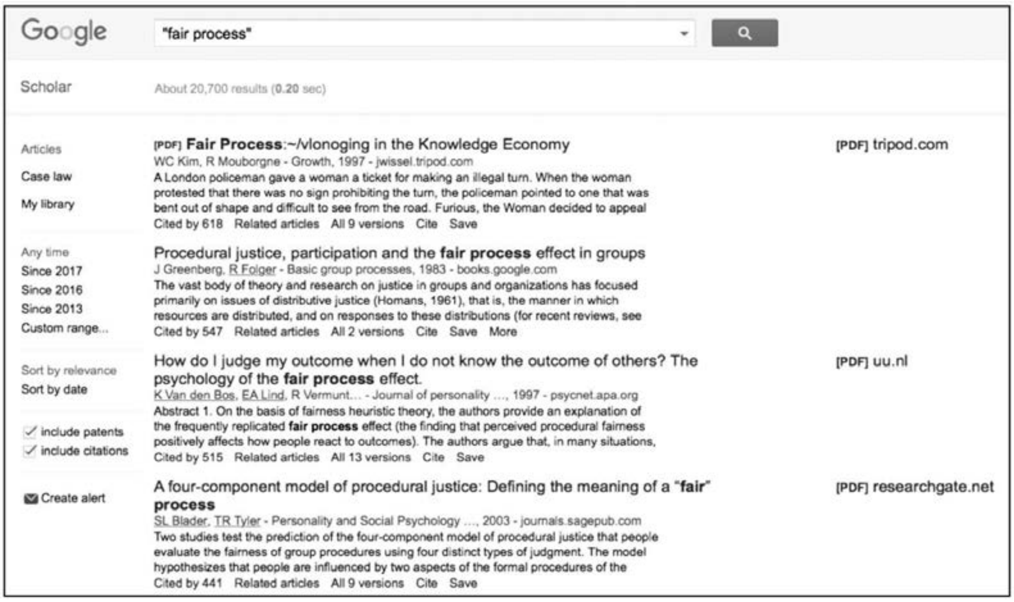
The question, therefore, is: *What is known in the research literature about the impact of a fair process on the way people perceive the outcomes of organizational change?*



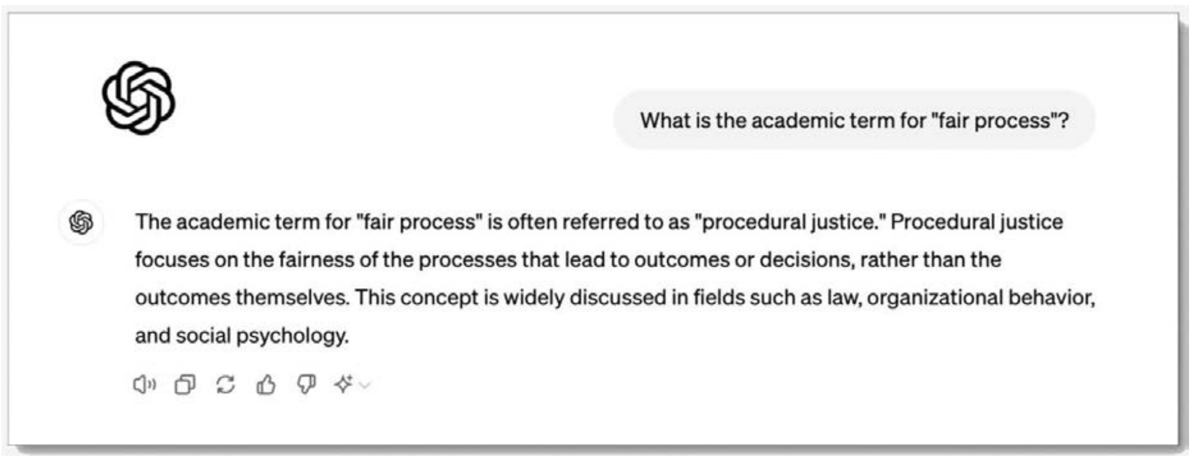
Using the search term "fair process" from the preceding example yields many results on Google, including some pages where alternative terms are mentioned. In addition, a search for alternative or related terms on the website Thesaurus.com yields terms such as honest, trustworthy, and objective. However, we are not sure whether these terms are also used in academia, so we need to check whether there are corresponding academic terms.

Google and AI Search Tools

The best place to find corresponding academic terms are Google Scholar and AI search tools such as ChatGPT and Consensus. If you search for “fair process” on Google Scholar, you will get many results. However, as you can see in the screenshot below, the results listed at the top suggest that there is a specific academic term for fair process: “procedural justice.” In addition, if you would skim through the first pages of the articles listed at the top, you would see that “procedural justice” is indeed a term (construct) that is widely used by academics and researchers.



AI tools such as ChatGPT are often the fastest way to find corresponding academic terms. As you can see in the screenshot below, when we asked ChatGPT for the academic term for "fair process," it correctly suggested "procedural justice." As you can see in the screenshot below, when we asked ChatGPT for the academic term for "fair process," it correctly suggested "procedural justice."



Consequently, if we searched only for “fair process” in a research database, we would find only a limited number of studies, while important, relevant studies using the academic term “procedural justice” would be missed.

Learn by doing 5.4

Imagine you would like to search for empirical studies on factors that enhance or inhibit employees' commitment to an organization. When you do a search on Google with the terms "commitment" and "organization" you will see that the corresponding (academic) construct is 'organizational commitment'

However, the academic construct 'organizational commitment' is an umbrella term, as there are three different types of organizational commitment, each with different and specific effects on outcomes.

Conduct a search on Google Scholar with the term "organizational commitment" and scroll through the results.

Determine which of the following terms are considered types of organizational commitment.
Select Yes or No.

A. Commitment to change
B. Continuance commitment
C. Affective commitment
D. Emotional commitment
E. Professional commitment
F. Normative commitment

Having a hard time finding the correct answer? Watch the video below to see how we searched.

https://tinyurl.com/mr32fzkr

Did I Get This 5.4

Imagine you would like to search for empirical studies on the effect of "360 degree feedback". Ask ChatGPT what corresponding academic terms are. Based on the outcome, determine which of the following terms are the most common used academic terms.

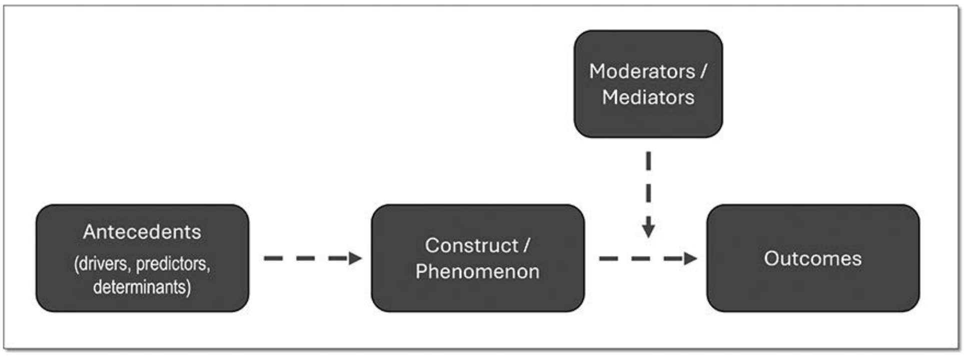
Select Yes or No.

A. Mult-source feedback
B. Full-circle feedback
C. Multi-rater feedback
D. Panoramic feedback

Aside: what is a 'construct'?

Some academic terms are referred to as 'constructs': an abstract idea or entity that can be observed and studied. In the natural sciences 'gravity', 'temperature', 'tectonic pressure', and 'global warming' are well known constructs. In the behavioral sciences constructs such as 'conscientiousness', 'self-esteem', and 'procedural justice' are widely used. A (psychological) construct is a label for a cluster of behaviors that serves as a hypothesized cause for an observed outcome. As such, constructs are the building blocks of scientific theories. The added value of constructs is that they represent a clearly defined concept that has the same meaning for different people. For example, the construct 'trust' is used by many behavioral scientists, but they all have a shared understanding of what it means: the willingness to be vulnerable to the actions of others.

When academics research a construct (see above) or phenomenon, they are often interested in several key questions: what the construct entails (its definition and how—if at all—it differs from related constructs), how it can be measured, what impact it has on certain outcomes, and what factors affect or predict it. In addition—especially when the impact on outcomes has been well established by previous research—they are interested in whether there are factors that influence how the construct affects those outcomes.



When academics study factors that predict a construct or phenomenon (e.g. absenteeism or turnover), they often use the term antecedent. Other terms commonly used are predictor or determinant. This means that if you want to find empirical studies on possible causes or drivers of, for example, absenteeism, you are more likely to find relevant results if you include search terms such as antecedent or determinant.

Likewise, when academics study factors that influence how a construct or phenomenon affects a particular outcome (e.g. factors that determine whether feedback leads to improved performance), they use the terms moderator and mediator. The exact meaning and function of moderators and mediators is explained in the next module (A Short Introduction to Science), but if you want to find empirical studies on contextual factors or mechanisms that may influence the impact of a construct or phenomenon, you may consider including the terms *moderator* or *mediator* in your search.

Conceptual Clarity

Whether or not you need to use alternative (academic) terms when searching for empirical studies depends on what is referred to as "**conceptual clarity**": how clearly and precisely a concept or construct is defined and how well it can be distinguished from related or overlapping concepts.

For example, the construct 'psychological safety' has strong conceptual clarity because it is clearly defined as "*… a shared belief among team members that it is safe for interpersonal risk taking—such as speaking up with ideas, questions, or concerns—without fear of embarrassment, rejection, or punishment.*" In addition, it can be measured in a reliable way and is distinguishable from related terms like 'interpersonal trust'. As a result, a single search using the term psychological safety is often sufficient.

The term '360-degree feedback', however, has less conceptual clarity. While there is a general understanding of what it entails, several alternative terms are used (e.g., *multi-source feedback, multi-rater feedback*), and it is not always clear what qualifies as '360-degree feedback'. As a result, finding relevant empirical studies requires multiple searches using a variety of related terms.

Searching becomes even more challenging when the topic or construct lacks conceptual clarity altogether—for instance, 'unethical behavior'. This term is not well-defined and can mean different things to different people. As a result, finding relevant empirical studies requires multiple searches using a variety of related terms, such as *misconduct, dishonesty, corruption, fraud, integrity, improper conduct, moral transgression*, and possibly more.

Learn by doing 5.5

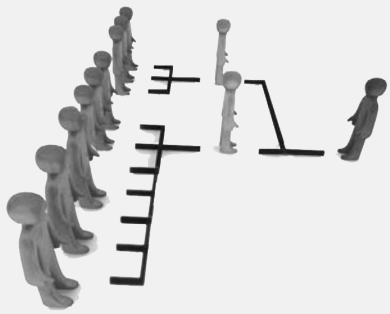
A manager in a large pharmaceutical company is responsible for several projects and change programs. Inspired by the idea of evidence-based management, she wonders whether there is scientific evidence on the success factors of change management. She therefore considers searching using the terms "change management" and "success factors."

How would you assess the conceptual clarity of the construct 'change management'?

Use ChatGPT to explore its definition and usage.

A. High
B. Moderate
C. Low

Did I Get This 5.5



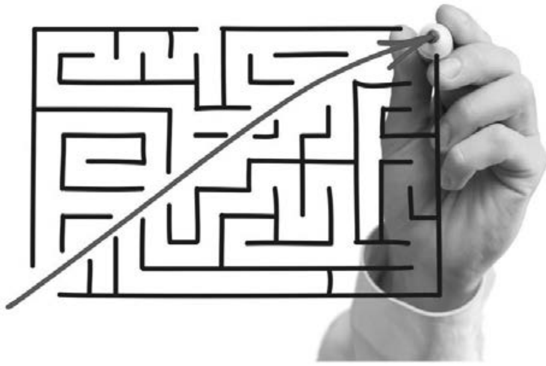
A British government agency wants to reduce costs and improve organizational efficiency. As part of this effort, it is reconsidering managers' 'span of control'—the total number of direct subordinates a manager oversees. The project manager leading this initiative wonders whether there is relevant scientific evidence on the topic and therefore considers conducting a search with the term "span of control."

How would you assess the conceptual clarity of the construct 'span of control'?

Use ChatGPT to explore its definition and usage.

A. High
B. Moderate
C. Low

Section 2: A Quick and Pragmatic Search

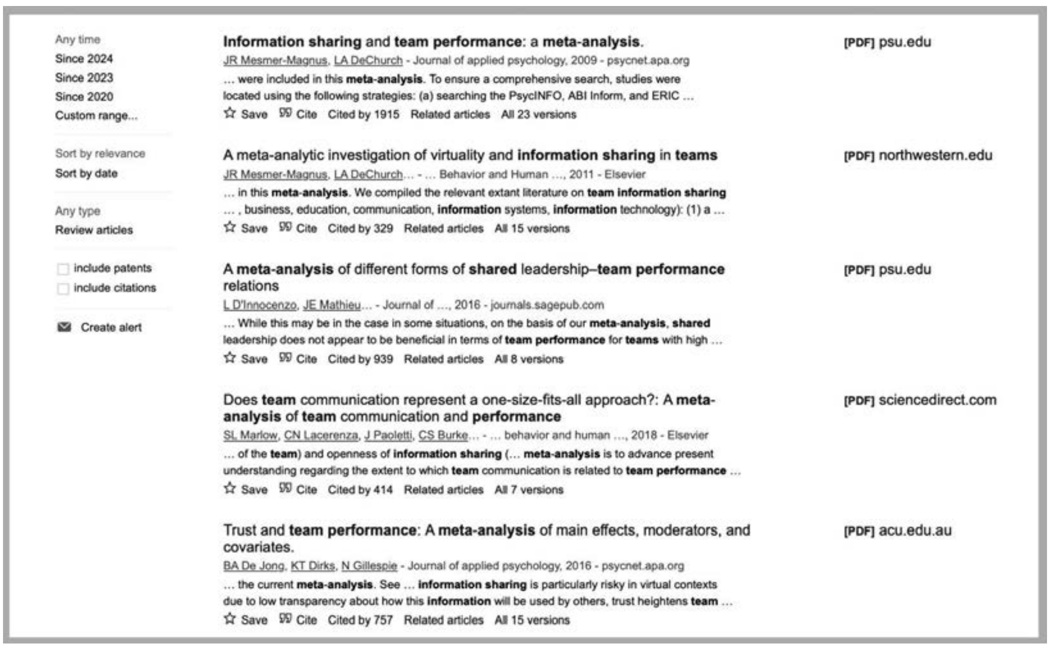


1. A quick and pragmatic search in Google Scholar

When we want to acquire evidence from the scientific literature, we have two options: a quick and pragmatic search in Google Scholar or AI tools, or a systematic and rigorous search across multiple research databases. As an evidence-based practitioner, you often just want to see whether there's relevant research regarding a bold claim someone makes (e.g. during a meeting or in a popular magazine) or on a topic that has your professional iterest. In those cases, a quick search in Google Scholar or Consensus often suffices.

This mode of searching takes a very pragmatic approach: Just type two of your PICOC terms into Google Scholar (usually the intervention and the outcome) and check whether there are relevant meta-analyses available. To ensure your search yields studies relevant to a specific context (the workplace in general or a specific domain such as healthcare) you can add another term.

In the example below we have searched with the terms 'information sharing', 'performance' and 'meta-analysis'. Because we are interested whether information sharing increases the performance of teams, we added the term 'teams'. As you can see in the screenshot below, Google Scholar finds several relevant meta-analyses, sufficient to get a general idea about the importance of information sharing in teams. If your search yields only a limited number of relevant papers, you can also try the option 'Related articles'.



Learn by doing 5.5

Read the following scenario:

Your neighbor tells you she just heard that the company she works for may be taken over by a large competitor. Although the board of directors has stated that no formal negotiations are underway, there are widespread rumors among employees that a deal is likely to be reached in the coming months. If the takeover happens, many positions may be made redundant, leaving your neighbor feeling very uncertain about her job. As a manager, you are professionally interested in whether job insecurity negatively affects employees' performance.

Conduct a quick and pragmatic search in Google Scholar using the search terms job insecurity, employee performance, and meta-analysis. Read the abstracts of the studies that seem most relevant. What do you find?

A. Job security does not affect employee performance
B. Job security positively affects employee performance
C. Job security negatively affects employee performance
D. Job insecurity can have both positive and negative effects on employee performance

Learn by doing 5.6

Read the following scenario:

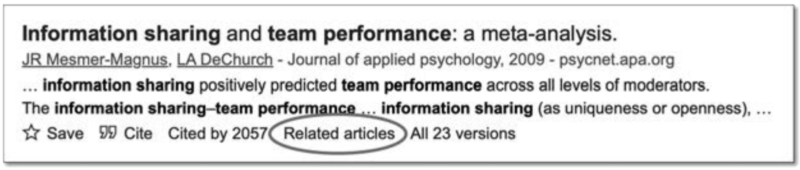
Your neighbor tells you she just heard that the company she works for may be taken over by a large competitor. Although the board of directors has stated that no formal negotiations are underway, there are widespread rumors among employees that a deal is likely to be reached in the coming months. If the takeover happens, many positions may be made redundant, leaving your neighbor feeling very uncertain about her job. As a manager, you are professionally interested in whether job insecurity negatively affects employees' performance.

Amlbsar _ osgai _lb np_ek_rga qc_paf gl Emmejc Qafmj_p sqgle rfc qc_paf rcpkq r_jj* fcgefr* n_w* c_plgleq* u_ec* glamkc _lb kcr_+_l_jwqgq, Pc_b rfc _`qrp_arq md rfc qrsbgcq rf_r qcck kmqr pcjct_lr, U f_r bm wms dgl b=

A. R_jj ncmnjc _pc n_gb kmpc
B. R_jj ncmnjc _pc n_gb jcqq
C. Rfcpc gq lm n_w bgddcpclac `crucc l r_jj _lb qfmpr ncmnjc
D. There are no meta-analyses on this topic

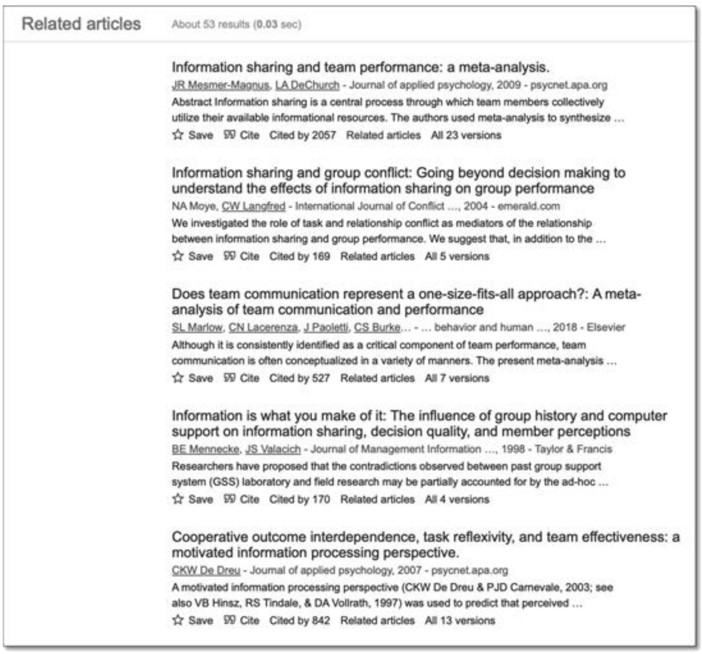
Snowballing

When your search in Google Scholar yields only a few relevant articles, you can use a technique known as ‘snowballing’ or ‘pearl growing’. This involves selecting the article you consider most relevant to your question and then 'snowballing' for additional studies on the same topic.



Information sharing and **team performance**: a meta-analysis.
JR Mesmer-Magnus, LA DeChurch - Journal of applied psychology, 2009 - psycnet.apa.org
… **information sharing** positively predicted **team performance** across all levels of moderators.
The **information sharing–team performance** … **information sharing** (as uniqueness or openness), …
Save Cite Cited by 2057 Related articles All 23 versions

To do this, click the ‘Related articles’ link located beneath the article in Google Scholar. This will display a list of articles that are similar in content to the one you selected.



Related articles About 53 results (0.03 sec)

Information sharing and team performance: a meta-analysis.
JR Mesmer-Magnus, LA DeChurch - Journal of applied psychology, 2009 - psycnet.apa.org
Abstract Information sharing is a central process through which team members collectively utilize their available informational resources. The authors used meta-analysis to synthesize …
Save Cite Cited by 2057 Related articles All 23 versions

Information sharing and group conflict: Going beyond decision making to understand the effects of information sharing on group performance
NA Moye, CW Langfred - International Journal of Conflict …, 2004 - emerald.com
We investigated the role of task and relationship conflict as mediators of the relationship between information sharing and group performance. We suggest that, in addition to the …
Save Cite Cited by 169 Related articles All 5 versions

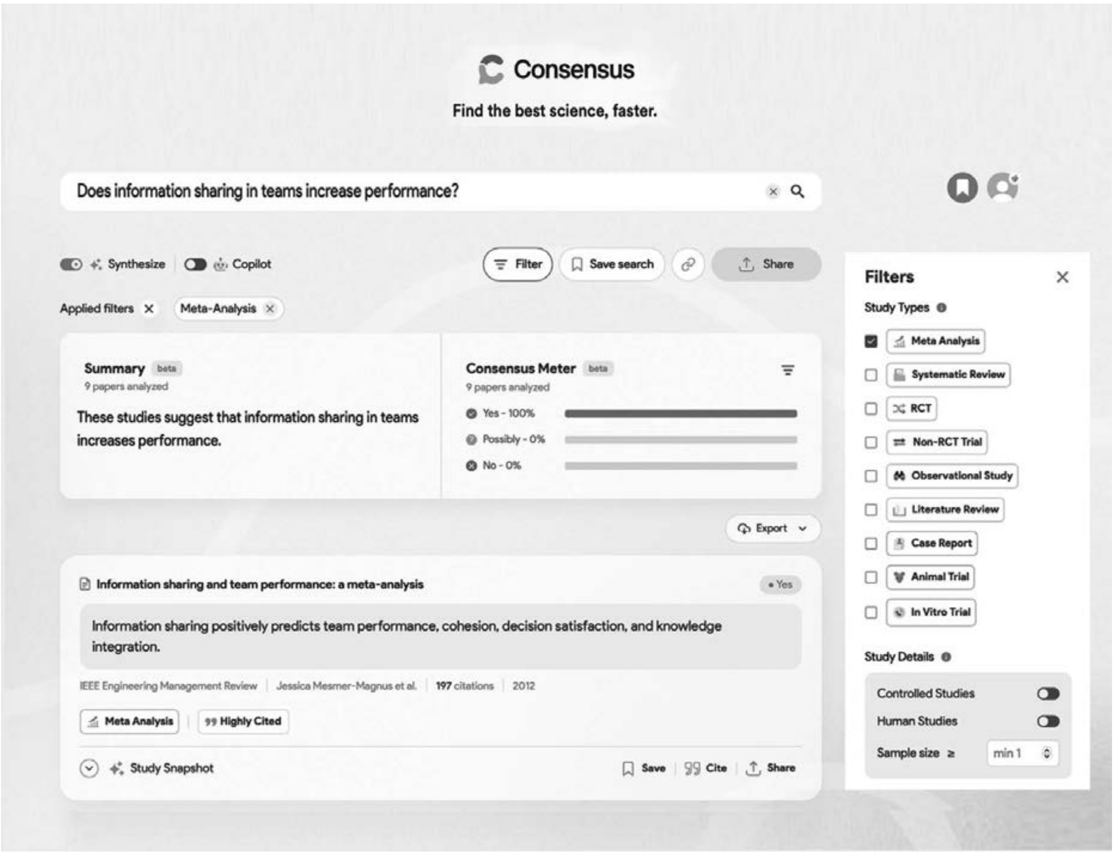
Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance
SL Marlow, CN Lacerenza, J Paoletti, CS Burke… - … behavior and human …, 2018 - Elsevier
Although it is consistently identified as a critical component of team performance, team communication is often conceptualized in a variety of manners. The present meta-analysis …
Save Cite Cited by 527 Related articles All 7 versions

Information is what you make of it: The influence of group history and computer support on information sharing, decision quality, and member perceptions
BE Mennecke, JS Valacich - Journal of Management Information …, 1998 - Taylor & Francis
Researchers have proposed that the contradictions observed between past group support system (GSS) laboratory and field research may be partially accounted for by the ad-hoc …
Save Cite Cited by 170 Related articles All 4 versions

Cooperative outcome interdependence, task reflexivity, and team effectiveness: a motivated information processing perspective.
CKW De Dreu - Journal of applied psychology, 2007 - psycnet.apa.org
A motivated information processing perspective (CKW De Dreu & PJD Carnevale, 2003; see also VB Hinsz, RS Tindale, & DA Vollrath, 1997) was used to predict that perceived …
Save Cite Cited by 842 Related articles All 13 versions

2. A quick in pragmatic search with AI tools

Nowadays, several AI search tools are available that specialize in finding research papers. Examples include Elicit, Perplexity, Semantic Scholar, Research Rabbit, and Consensus. Most of these search tools work a little differently from Google Scholar. For example, Consensus does not work with search terms but with research questions. As you can see in the screenshot below, this tool finds nine meta-analyses that seem relevant to our question "Does information sharing in teams improve performance?". In addition, Consensus offers extra functionalities, such as filtering by study design, sample size, and a 'consensus meter' that classifies the results based on whether they provide evidence in support of a claim or question.



Did I Get This 5.6

Read the following scenario:

You're attending a meeting of the executive team at a large manufacturing company. Over the past year, as part of a new performance management system, annual performance assessment meetings have been introduced—each supervisor and manager is now expected to meet with every subordinate to discuss their performance. During the meeting, the HR director expresses concern that most managers are too lenient. He argues that they should also provide negative feedback, especially when an employee's performance falls below expectations. Apparently, the HR director assumes that negative feedback has a positive effect on performance.

Conduct a quick and pragmatic search in Consensus to see whether scientific evidence supports the HR director's assumption that negative feedback improves employee performance.

What do you find?

A. Negative feedback has a positive effect on employee performance
B. Negative feedback can have both positive and negative effects on employee performance
C. Negative feedback has a negative effect on employee performance

Keep in mind, however, that a quick and dirty search in Google Scholar or Consensus has many limitations. First, neither Google Scholar nor Consensus have access to all research (a lot of research is behind pay walls, especially studies published in top academic journals), so they don't provide a fully comprehensive look into all studies regarding your question. As a result, important studies may be missed.

Second, an evidence-based search is supposed to be systematic, transparent, and reproducible. The fact is that no one knows exactly how Google Scholar and AI tools like Consensus search and select papers. There are several examples of AI tools being biased or producing faulty results, so be sure to do your due diligence.

A quick search in Google Scholar or Consensus is a great starting point and often sufficient for quickly checking a claim or assumption, but sometimes a more rigorous search is needed.

When an important decision has to be made that involves spending significant resources and that will affect a large number of people, a quick search as described above is often insufficient. In these cases, we need to go for the 'gold standard': a highly systematic, reproducible, and thorough search to identify the most relevant high-quality empirical studies in research databases such as ABI/INFORM, Business Source Elite, and PsycINFO.

Unfortunately, the 'gold standard' is time-consuming and requires specific skills and a lot of practice. In the next part of this module, we will explain how to conduct a systematic and 'deep' search in research databases and provide the opportunity to further develop your search skills.

Did I Get This? 5.7

Read the following scenario:

During an executive meeting, the Director General of the Dutch Ministry of Infrastructure and Water Management says she believes that the Ministry's performance and efficiency can be improved by implementing "autonomous teams." According to the Director, these teams empower civil servants to plan their own work, access resources and information directly, and make decisions without constant managerial oversight. She strongly believes this approach will boost both performance and efficiency, but asks you to examine whether this is supported by scientific evidence.

In this scenario, what is the most appropriate mode of searching?

A. A quick and pragmatic search in Google Scholar or AI search tools
B. A systematic and rigorous search in multiple research databases

Did I Get This 5.8

Read the following scenario:

During an executive meeting, the Director General of the Dutch Ministry of Infrastructure and Water Management says she believes that the Ministry's performance and efficiency can be improved by implementing "autonomous teams." According to the Director, these teams empower civil servants to plan their own work, access resources and information directly, and make decisions without constant managerial oversight. She strongly believes this approach will boost both performance and efficiency, but asks you to examine whether this is supported by scientific evidence.

In this scenario, what is the most appropriate mode of searching?

A. A quick and pragmatic search in Google Scholar or AI search tools

B. A systematic and rigorous search in multiple research databases

Did I Get This 5.9

Read the following scenario:

Over lunch, your colleague mentions an article he read in a popular business magazine, which claims that an increasing number of organizations are adopting "autonomous teams." According to the article, these teams—where employees plan their own work, access resources directly, and make decisions without constant oversight—boost financial performance by an average of 15%. Convinced that autonomous teams could benefit your organization as well, your colleague asks whether this claim is supported by scientific evidence.

In this scenario, what is the most appropriate mode of searching?

A. A quick and pragmatic search in Google Scholar or AI search tools

B. A systematic and rigorous search in multiple research databases

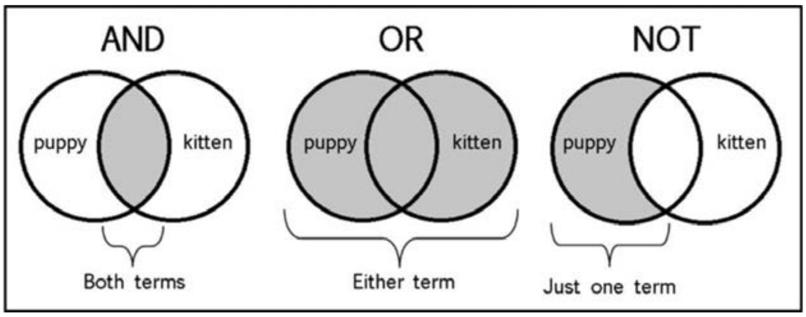
Section 3: A Systematic and Rigorous Search

Searching With Boolean Operators

In contrast to Google Scholar and Consensus, research databases such as ABI/INFORM, Business Source Elite, and PsycINFO provide several options by which to specify your search. As you can see in the following screenshot, the interface of a research database allows you to search for keywords in the title and/or the abstract. You simply enter your search terms in the search field and click the dropdown menu on the right. In addition, you can use multiple keywords. Research databases make use of Boolean operators (AND, OR, and NOT), which allow you to search for keywords in different combinations.

The Boolean operator **OR** increases the number of results you retrieve and is used to combine synonyms or related terms to make your results more comprehensive. For example, entering “360-degree feedback” [keyword 1], OR [in the dropdown], and “multisource feedback” [keyword 2] returns articles that mention *either* of these topics in the title or abstract.

Conversely, **AND** reduces the number of results you retrieve and is used to combine PICOC terms or methodological filters to make your results more relevant. For example, entering “cultural diversity” AND “performance” returns articles that mention *both* topics in the title or abstract.

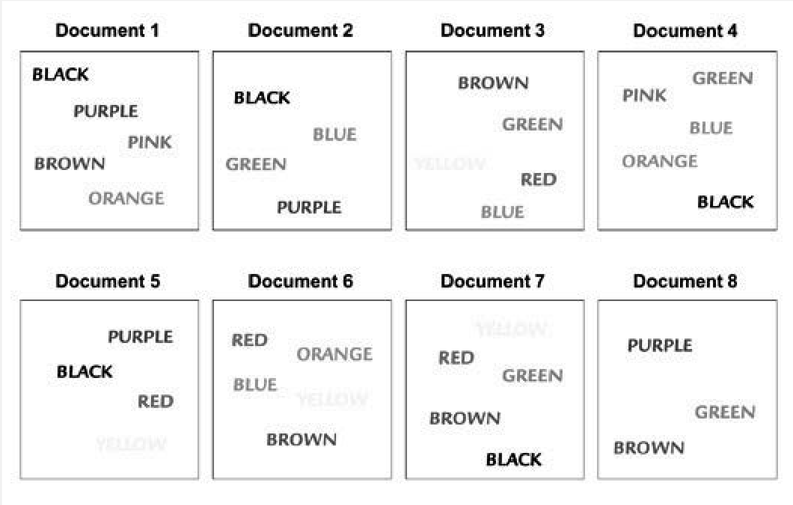


The Boolean operator **NOT** reduces the number of results you retrieve by excluding a specific term. For example, entering “diversity” NOT “cultural” returns articles that mention diversity in the title or abstract but removes any articles that mention cultural.

Keep in mind that OR broadens your search, whereas AND narrows your search

Learn by doing 5.7

Imagine you have a database with eight documents in it, and you are going to do a search. Here is a visual illustration of the database and the documents it contains.



1. You are searching with the terms ***red* AND *blue***. Which documents will your search yield?
   a. Document 3
   b. Documents 3 and 6
   c. Documents 3, 5, 6, and 7
   d. Documents 2, 3, 4, 5, 6, and 7
2. You are searching with the terms ***pink* OR *purple***. Which documents will your search yield?
   a. Document 1
   b. Documents 1 and 4
   c. Documents 1, 2, and 5
   d. Documents 1, 2, 4, 5, and 8

Did I get this 5.7

You are searching with the terms ***brown* AND *green***. Which documents will your search yield?

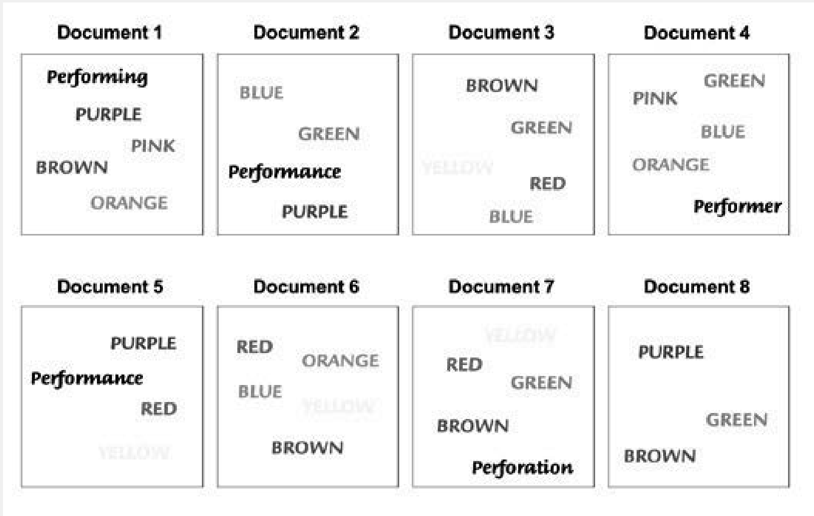
a. Document 3
b. Document 5
c. Documents 3, 7 ,and 8
d. Documents 1, 2, 3, 4, 6, 7, and 8

TIP: Truncation*

Most research databases allow you to use the truncation symbol * (the asterisk) for finding singular and plural forms of words and variant endings. For example, typing “work*” into the search field returns articles containing any of the following words in the title or abstract: work*s*, work*er*, work*ers*, work*ing*, work*force*, and work*place*.

Did I get this 5.10

Imagine you have a database with eight documents in it, and you are going to do a search. Here is a visual illustration of the database and the documents it contains.



You are searching with the terms ***green*** **AND** ***perform****. Which documents will your search yield?

a. Documents 2 and 4

b. Documents 3 and 6

c. Documents 2, 4, and 7

d. Documents 1, 2, 3, 4, 5, 7, and 8

Pre-testing your search terms

After you have identified alternative terms and corresponding academic terms, it is important to pretest your search terms: see which of the terms you have identified yield the most relevant results in the research databases you have selected. When you conduct a quick and pragmatic search in Google Scholar or Consensus, pre-testing search terms is often not necessary, as these tools will automatically search using alternative terms. However, a search in a research database will only yield results for the search term you have entered.

In general, a pretest in just one research database will already give you a good impression of which search terms yield the most relevant results (sensitivity) while minimizing the number of irrelevant results (specificity).

Example 1

In 2015, McKinsey & Company, a prestigious international consulting firm, published a research report, "Why Diversity Matters," in which it claimed that companies with an ethnically diverse workforce outperformed nondiverse companies. Based on this report, the HR director of a Danish company specializing in children's furniture considered setting up a project to increase the diversity of its workforce. Before a decision was made, however, she first wanted to find out what was known in the scientific literature about the effect of ethnical diversity on workplace performance. After she had formulated her PICOC,

P: Manufacturing workers
I: Ethnic diversity
C: No diversity
O: Performance
C: Danish company specializing in the production of children's furniture

... she decided that 'ethnic diversity' and 'performance' were the two most important PICOC terms. After a search on Google and Google Scholar, she found several related terms, such as 'cultural diversity', 'demographic diversity', 'heterogeneity', and 'minority'. When she pretested her search terms in the research database Business Source Premier, she got the following results:

Search term	Results	Search term	Results
TI diversity	4,880	TI "demographic* divers*"	65
AB diversity	21,334	AB "demographic* divers*"	1,303
TI "ethnic* divers*"	224	TI "heterogen* AND work*"	131
AB "ethnic* divers*"	1,832	AB "heterogen* AND organization*"	20
TI "cultur* divers*"	458	TI minority	2,386
AB "cultur* divers*"	5,169	TI (minorit* AND work*)	75

When she skimmed through the titles and the abstracts of the articles found, she noticed that a search with a combination of the terms 'divers', 'ethnic*', 'culture*', and 'demograph*' in the title yielded the most relevant results, while other terms resulted in many articles that were irrelevant.*

Example 2

In recent years, events at several banks have led regulators to place a far greater emphasis on staff conduct and ethical workplace behavior. The chief risk officer of a large international banking firm in the United Kingdom therefore considered setting up a project to decrease the risk of staff misconduct. However, he decided to first find out what was known in the scientific literature about the drivers of unethical workplace behavior and misconduct. After he had formulated his PICOC,

P: Employees at a bank
I: Factors, drivers
C: Not specific
O: Unethical behavior, misconduct
C: Global banking firm in the United Kingdom

... he determined that unethical behavior and misconduct were the two most important PICOC terms. After conducting a search on Google and Google Scholar, he found a large number of alternative and related terms, such as 'integrity', 'fraud', 'rule breaking', 'compliance', and 'ethical climate'. When he pretested his search terms using the research database Business Source Premier, he got the following results:

Search term	Results	Search term	Results
TI unethical	377	TI integrity	1,289
TI "unethical behavior"	100	TI (integrity AND work*)	27
AB "unethical behavior"	525	TI fraud*	2,801
TI (unethical AND work*)	21	TI (fraud* AND work*)	28
TI "unethical workplace behavior"	0	TI (break* rule*)	14
AB "unethical workplace behavior"	0	TI (compliance AND rule*)	107
TI misconduct	327	TI (ethic* AND organization*)	712
AB misconduct	1,371	TI (ethic* AND work*)	919
TI (misconduct AND work*)	44	TI dishonest*	162
TI "ethical climate"	136	TI (dishonest* AND work*)	8
AB "ethical climate"	282	TI (compliance AND work*)	4

When he skimmed through the titles and the abstracts of the articles found, he noticed that a search with a combination of the terms 'unethical behavior', 'misconduct', and 'ethical climate' yielded the most relevant results.

Learn by doing 5.8

Go to Business Source Elite.

Imagine you would like to find studies on characteristics of entrepreneurs (people who independently own and actively manage a business). Go to Business Source Elite and check which of the following terms yields **the most relevant** results.

NOTE:

Search in the **title** and make sure you use quotation marks for search terms of more than one word (e.g., “business owner”).

1. entrepreneur*
2. enterpriser*
3. industrialist*
4. business owner*
5. tycoon
6. founder CEO*

Having a hard time finding the cor ect answer? Watch the video below.

https://youtu.be/xsglO194S4M

Did I get this 5.11.

Go to Business Source Elite.

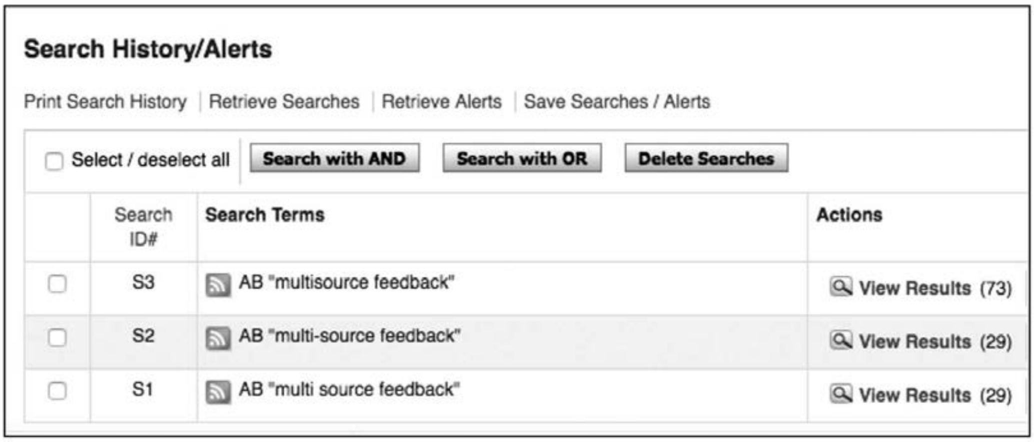
Imagine you would like to find studies on the effect of financial bonuses on the performance of executives and CEOs. Go to Business Source Elite and check which of the following terms in the title yields the most relevant results.

1. Financial reward*
2. Financial fee*
3. Financial incentive*
4. Pay for performance plans
5. Performance reward*
6. Variable pay

Pretesting is important not only for relevance but also for spelling!

When it comes to spelling, even for native speakers, English is a difficult language. For example, should we search for "multisource feedback," "multi-source feedback," or "multi source feedback"? As you can see below, the terms "multi source feedback" and "multi-source feedback" both yield 29 results, whereas the term "multisource feedback" (the most common spelling) yields 73 results.

Consequently, it makes sense to always test whether a different spelling yields different results.



Search History/Alerts

Print Search History | Retrieve Searches | Retrieve Alerts | Save Searches / Alerts

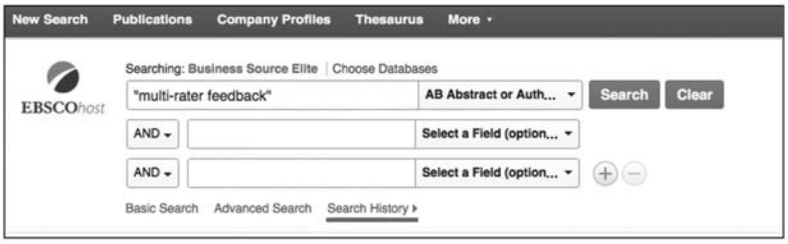
Select / deselect all | Search with AND | Search with OR | Delete Searches

	Search ID#	Search Terms	Actions
	S3	AB "multisource feedback"	View Results (73)
	S2	AB "multi-source feedback"	View Results (29)
	S1	AB "multi source feedback"	View Results (29)

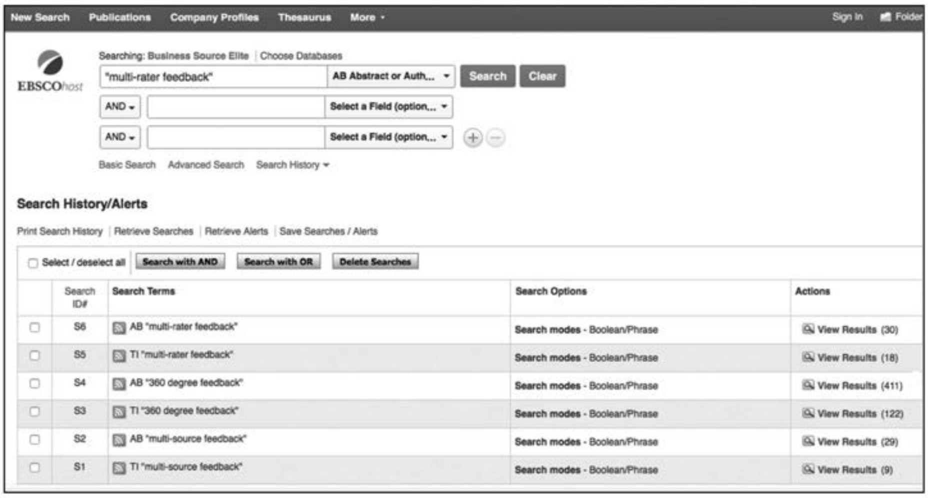
Note that spelling is less important when searching in Google Scholar or with AI search tools such as Consensus, as these tools often account for spelling variations.

Using The History Function to Combine Searches

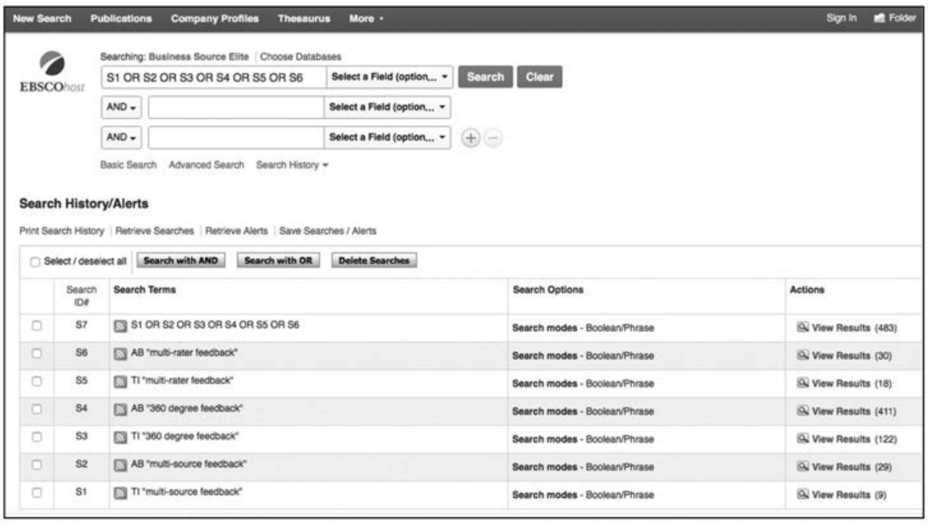
Another great feature that all research databases have is the search history. All searches performed during your search session are available from the Search History screen. You can open the screen by clicking the **Search History** link under the search fields



In this screenshot, you can see that we have searched for the keywords “multisource feedback,” “multirater feedback,” and “360 degree feedback” in the title (TI) or the abstract (AB).



We can now combine all our search queries with OR to create one big bucket with all articles that may be relevant. We can do that in two ways: we can either select all search queries by clicking the checkboxes and then click Search with OR, or we can just type in “S1 OR S2 OR . . .” in the search field. In the screenshot below, you can see the latter option returns about 480 results.



Did I get this 5.12

As you can see in the screenshot above, when the individual searches are combined with OR, the total number of results is 483. However, when we add the results of each separate search, we get a much larger number (9 + 29 + 122 + 411 + 18 + 30 = 619 results.

How is this possible?

a. This is probably a calculation error.
b. Some articles use more than one search term in the title or the abstract.
c. Only articles that use more than one search term in the title or the abstract are included.
d. Articles that use more than one search term in the title or the abstract are not included

Learn by doing 5.9

Go to Business Source Elite.

Imagine you want to find empirical studies on the differences between entrepreneurs and managers. Your search for alternative terms for entrepreneurs has yielded the following terms: “**business owner**” and ‘**founder**’. Search in Business Source Elite for ‘**entrepreneur**’ and these three alternative terms in the title, and then use the history function to combine the outcome with OR.

How many peer-reviewed articles does your search yield?

a. Less than 300 results
b. Between 300 and 400 results
c. Between 3,000 and 4,000 results
d. More than 4,000 results

Having a hard time finding the correct answer? Watch the video below to see how we searched.

https://youtu.be/JbsmRfsa9BY

Did I get this 5.13

Imagine you want to find empirical studies on the effect of cultural diversity on team performance. Your search for alternative terms for “**cultural diversity**” has yielded the following two terms: “**ethnic diversity**” and “**racial diversity**”.

Search for “cultural diversity” and these two alternative terms in the title, and then use the history function to combine the outcome with OR.

How many peer-reviewed articles does your search yield?

a. Less than 400 results
b. Between 400 and 1000 results
c. Between 4,000 and 5,000 results
d. More than 5,000 results

Having a hard time finding the correct answer? Watch the video below to see how we searched.

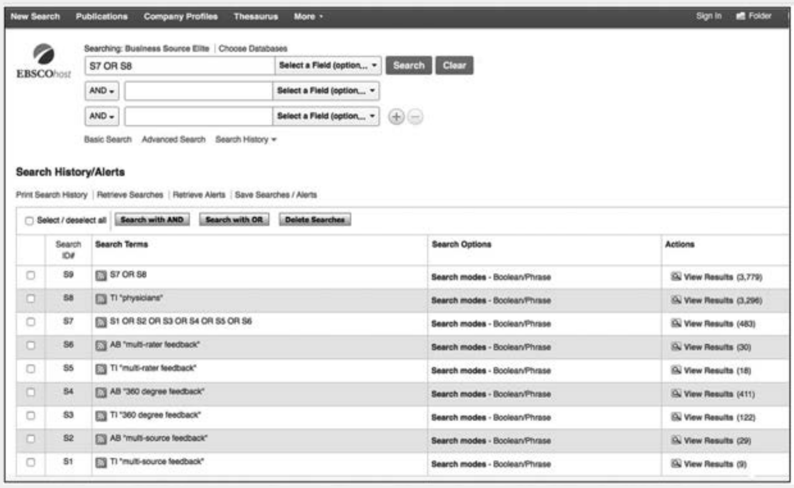
https://youtu.be/ToCfHIsVyI4

Why do I have more/less results than in the examples?

The number of academic papers in a research database changes on a daily basis. Every day a large number of new papers are published. In addition, certain journals may no longer be included, or new journals may be added to the database. Consequently, the number of your search results may somewhat differ from the those in the examples and screenshots.

Did I get this 5.14

In this screenshot, you can see that the researcher has combined multiple searches with 360-degree feedback and related terms. Because the researcher is interested in whether 360-degree feedback has a positive effect on the performance of doctors they have added the PICOC term physicians.
As you can see, the number of results has increased from 483 to 3,779. What went wrong?



a. All articles that use more than one search term in the title or the abstract are also included.
b. The researcher should have added more PICOC terms to limit the results.
c. The researcher should have used AND instead of OR.
d. The researcher probably clicked the Search button twice.

Searching for Meta-analyses

Searching in Research Databases: Systematic and reproducible

As you noticed, we have dedicated quite some time to identifying the right search terms. We explained why this is important: garbage in, garbage out – good search terms result in good outputs, and unclear, ambiguous, or incorrect search terms result in bad outputs. Whereas the phase of finding the right sea ch terms is characterized by a trial-and-error approach and a rather associative and explorative search process, the search for empirical studies in a research database is highly systematic. In general, the next six steps are typically followed:

Step 1. Conduct a search with your (pretested) search terms and combine the outcome with OR.

Step 2. Filter the combined outcome for meta-analyses and/or systematic reviews.

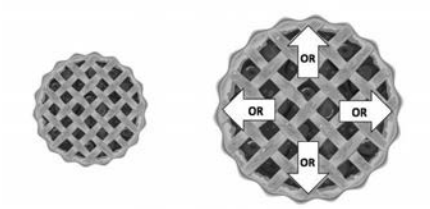
Step 3. Filter the combined outcome for high-quality primary studies.

Step 4. *If necessary*, filter the combined outcome for low-quality primary studies

Step 5. *If necessary*, limit the number of results by adding a second or third PICOC term.

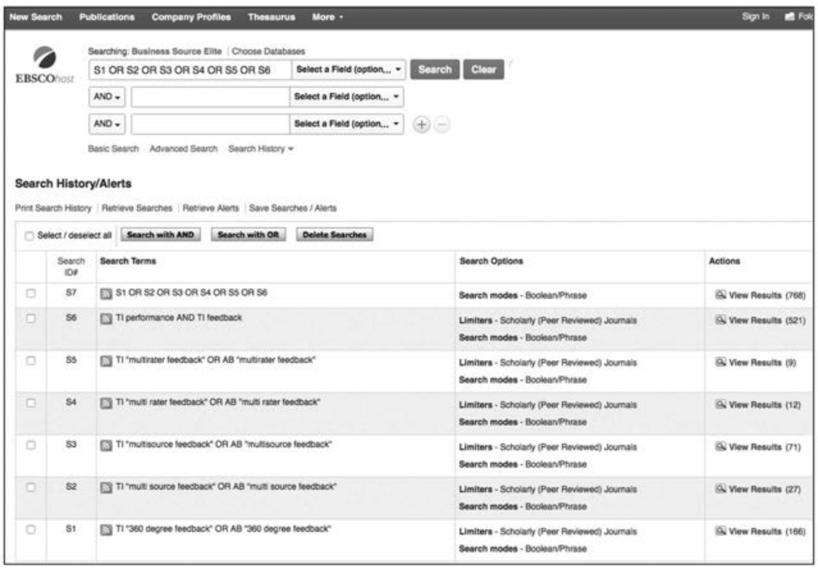
Step 6. Screen the articles found for relevance.

Step 1. Enlarge the pie: Conduct a search with your (pretested) search terms and combine the outcome with OR.



The first step is to enlarge the "pie": conduct a search with each of your (pretested, most relevant) search terms and combine the outcomes with the Boolean operator OR. This pie will be the basis for the next steps of your search. In the screenshot below, you can see how this is done using the example regarding the effectiveness of 360-degree feedback.

Note that we have conducted a search for both *multi source feedback* and *multisource feedback* because we know from our pretest that these terms yield different (but relevant) results. We do the same with *multi rater* and *multirater*. Because we found that *performance feedback* is an underlying construct, we have also conducted a search with the terms *performance* and *feedback* in the title. Finally, we have combined the six search queries with OR, which resulted in a large pie with 768 peer-reviewed papers.



Did I get this 5.15

Look at the following screenshot. A person has conducted (almost the same search as above but instead ended with 116,507 results.

What went wrong?

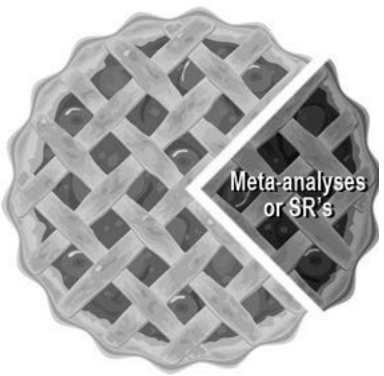
Search History/Alerts

Print Search History | Retrieve Searches | Retrieve Alerts | Save Searches / Alerts

Select / deselect all | Search with AND | Search with OR | Delete Searches

	Search ID#	Search Terms	Search Options	Actions
	S7	S1 OR S2 OR S3 OR S4 OR S5 OR S6	Search modes - Boolean/Phrase	View Results (116,507)
	S6	TI performance OR TI feedback	Search modes - Boolean/Phrase	View Results (116,507)
	S5	TI "multirater feedback" OR AB "multirater feedback"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (9)
	S4	TI "multi rater feedback" OR AB "multi rater feedback"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (12)
	S3	TI "multisource feedback" OR AB "multisource feedback"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (71)
	S2	TI "multi source feedback" OR AB "multi source feedback"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (27)
	S1	TI "360 degree feedback" OR AB "360 degree feedback"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (166)

Step 2: Cut the pie: Filter the combined outcome for meta-analyses



A systematic review or meta-analysis is a research paper in which the authors systematically searched for and summarized the findings of relevant studies on the same topic. In fact, meta-analyses or systematic reviews are often the best available scientific evidence on a topic. (1) Both types of research paper often provide the best available scientific evidence on a topic. You will learn more about meta-analyses in module 6 and 7.

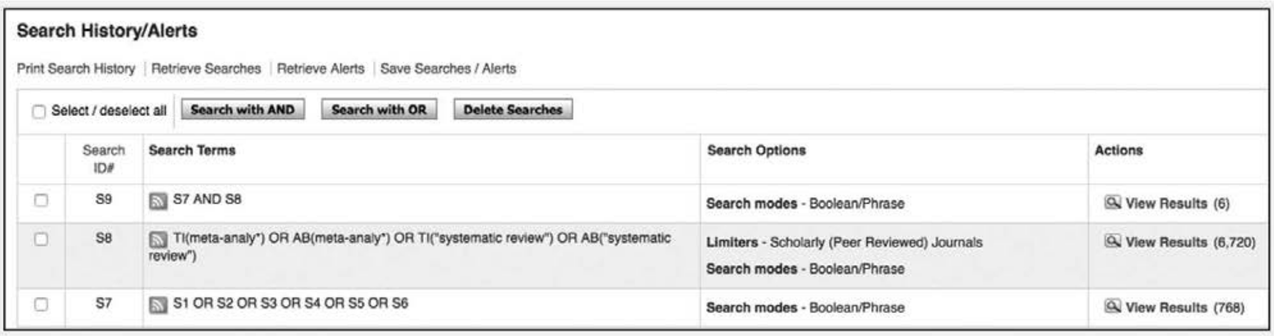
Thus, if we find a meta-analysis, we can save ourselves a lot of time and effort. However, it is too time consuming to read 768 abstracts to find out whether an article concerns a meta-analysis or systematic review, so we have to find a clever way to filter them out.

We therefore apply a methodological search filter that was developed by the Center for Evidence-Based Management (CEBMa). You can copy and paste the filter directly into the search box of the database. Do not select a specific field (e.g., title or subject term) in the dropdown list, as doing so will negatively affect your search.

Filter to Identify Meta-analyses and Systematic Reviews

TI(meta-analy*) OR AB(meta-analy*) OR TI("systematic review") OR AB("systematic review")

When we run a search with this filter and combine the outcome in the search history with the set of 768 papers by using AND, we get six results.



Search History/Alerts

Print Search History | Retrieve Searches | Retrieve Alerts | Save Searches / Alerts

Select / deselect all | Search with AND | Search with OR | Delete Searches

	Search ID#	Search Terms	Search Options	Actions
☐	S9	S7 AND S8	**Search modes** - Boolean/Phrase	View Results (6)
☐	S8	TI(meta-analy*) OR AB(meta-analy*) OR TI("systematic review") OR AB("systematic review")	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (6,720)
☐	S7	S1 OR S2 OR S3 OR S4 OR S5 OR S6	**Search modes** - Boolean/Phrase	View Results (768)

Learn by doing 5.10

Go to Business Source Elite.

Conduct a search with

1. “feedback” in the Title
2. “330 degree feedback” in the Title OR Abstract
3. “multisource feedback” in the Title OR Abstract
4. “multirater feedback” in the Title OR Abstract

Use the history function to combine the outcome with OR. Then apply the filter to identify meta-analyses.

Your final result should be 25 to 35 papers (as explained, this number may change, as studies are sometimes added to or removed from the database).

If your final result is much higher (or zero), check the video below to see where you may have made a mistake.

https://youtu.be/wi3FHKivlHI

Look at the titles of the papers in your final set and determine which of the following titles are included.

a) “Does performance improve following multisource feedback? A theoretical model ...”
b) “Feedback effectiveness: Can 360-degree appraisals be improved?”
c) “An objective review of the effectiveness and essential characteristics of performance feedback ...”
d) “Leveraging multirater feedback to facilitate successful behavioral change”

Learn by doing 5.11

Go to Business Source Elite.

Imagine you want to find empirical studies on the effect of cultural diversity on team performance. Your search for alternative terms for *cultural diversity* **has yielded two terms:** *ethnic diversity* **and** *racial diversity*. **Search for studies in Business Source Elite with cultural diversity and these two alternative terms in the title, and use the history function to combine the outcome with OR. Then apply the filter to identify meta-analyses.**

How many peer-reviewed articles does your search yield?

a. About 1 to 10 results
b. About 10 to 20 results
c. About 20 to 50 results
d. More than 50 results

Having a hard time finding the correct answer? Watch the video below.

https://youtu.be/ur7op9a3SUk

Learn by doing 5.12

Look at the titles of the papers in your final set and determine which of the following titles are included.

Search ID#	Search Terms	Search Options	Actions
S6	S4 AND S5	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (6)
S5	TI(meta-analy*) OR AB(meta-analy*) OR TI("systematic review") OR AB("systematic review")	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (12,263)
S4	S1 OR S2 OR S3	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (601)
S3	TI "racial diversity"	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (57)
S2	TI "ethnic diversity"	Limiters - Peer Reviewed	View Results (215)

a) "Framing the effects of multinational cultural diversity on virtual team processes"
b) "Ethnic diversity and creativity in teams: A review of 30 years of research"
c) "Cultural diversity in the workplace: The state of the field
d) "Unraveling the effects of cultural diversity in teams: A meta-analysis of research ..."
e) "Racial diversity, age diversity, interdependence, and team performance"

Did I get this 5.16

Go to Business Source Elite.

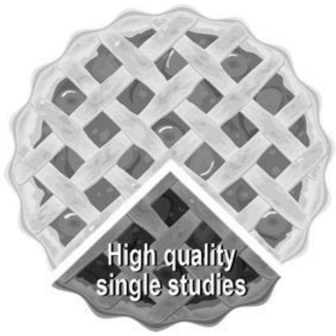
Imagine you want to find empirical studies on virtual teams. Your search for alternative terms for *virtual teams* **yields only one term:** *distributed teams***. Search for studies in Business Source Elite with these two terms in the title and the abstract, and use the history function to combine the outcome with OR. Then apply the filter to identify meta-analyses.**

Look at the titles of the papers in your final set and determine which of the following titles are included.

a) "Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness ..."
b) "Virtual teams: A review of current literature and directions for future research"
c) "Managing virtual teams: A review of current empirical research"
d) "Cultural influences on early trust development in virtual teams"

Searching for primary studies

Step 3. Search for high-quality single studies.



If you found one or more relevant meta-analyses, you may already have sufficient scientific evidence to answer your question. Unfortunately, for many topics, meta-analyses and systematic reviews are not available. In that case, your search will not yield any (or will yield only irrelevant) articles, and you will have to look for high-quality primary studies – that is, controlled and/or longitudinal studies. To find high-quality single studies, first, enlarge the pie: conduct a search with your (pretested) search terms and combine the outcome with OR. This search query is probably still present in the search history. Then cut the pie by applying a second methodological search filter to find controlled and/or longitudinal studies. Again, you can copy the filter below and paste it directly into the search box of the database.

Filter to Identify Controlled and/or Longitudinal Studies

TI(experiment* OR controlled OR longitudinal OR randomized OR quasi) OR AB(experiment* OR “controlled stud*” OR “controlled trial” OR “control group” OR “control variable” OR “comparison group” OR “comparative stud*” OR quasi OR longitudinal OR randomized OR randomly OR laboratory OR “before and after stud*” OR “pretest post*” OR “time series” OR “case control” OR “case cohort” OR “cohort stud*” OR “prospective stud*”)

Please keep in mind that the sensitivity (true positive rate) and specificity (true negative rate) of this filter i limited. As a consequence, your search results may still contain low-quality or theoretical studies, while some high-quality studies may be missed.

In the following example, we have applied this filter to our search for empirical studies on *360-degree feedback*. In the video below you can see our search yielded 13 studies.

https://youtu.be/DGwyY_-vobU

Learn by doing 5.13

Go to Business Source Elite.

Conduct the same search as shown in the video above. Your final result should be about 10 to 20 papers (as explained, this number may vary).

Look at the titles of the papers in your final set and determine which of the following titles are included.

a) “Effects of multisource feedback and a feedback facilitator ...”
b) “Does performance improve following multisource feedback? A theoretical model, meta-analysis ...”
c) “When the purpose of using multi-rater feedback is behavior change”
d) “Personality and multisource feedback improvement: A longitudinal investigation”

Did I get this 5.17

Go to Business Source Elite.

Imagine you want to find empirical studies on virtual teams. Do a quick search for studies with the term “**virtual teams**” in the title. Then apply the filter to identify longitudinal and/or controlled studies. You should have about 40 to 60 results.

Look at the first 20 titles of the papers in results list and determine which of the following titles are included.

a) “In-group/out-group effects in distributed teams: A meta-analysis and research synthesis”
b) “Innovation and communication media in virtual teams – An experimental study.”
c) “Out of sight, out of sync: Understanding conflict in distributed teams
d) “Trust and the unintended effects of behavior control in virtual teams”

Step 4. Search for low-quality single studies.

For some topics, even controlled and/or longitudinal studies are not available. In that case, we are left with no option other than to search for low-quality studies. In the step 3 example (360-degree feedback), we could go through all the titles and abstracts of all of the articles we have retrieved with the combined search query, but most of these articles may concern nonempirical studies such as essays and theoretical papers. For this reason, we conduct a final search query that selects only articles that mention the word “study” in the abstract. As you can see in the example below (S6 and S7), this yields 58 results.

This method, however, is not very reliable and should be used only when no (or a very limited number of) high-quality studies are available.

Search ID#	Search Terms	Search Options	Actions
S6	S4 AND S5	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (58)
S5	AB study	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (964,806)
S4	S1 OR S2 OR S3	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (124)
S3	TI "multirater feedback" OR AB "multirater feedback"	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (11)
S2	TI "multisource feedback" OR AB "multisource feedback"	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (77)
S1	TI "360 degree feedback" OR AB "360 degree feedback"	Limiters - Peer Reviewed Expanders - Apply equivalent subjects Search modes - Boolean/Phrase	View Results (36

Did I get this 5.18

Go to Business Source Elite.

In one of the previous sections, we explained that many organizations today have embraced Agile, a new method for managing projects. An essential part of Agile is the daily scrum: a short stand-up meeting at the beginning of the day during which team members share information and speak up about any problems that might prevent project completion. Imagine you would like to know whether any empirical studies are available on this topic yet. Because this topic is quite novel, there probably will be no meta-analyses (yet and only a limited number of high-quality studies. Therefore, conduct a search with *scrum* in the title and then conduct a search to identify studies that use the term *study* in the abstract.

How many studies did you find?

a) 100 – 500
b) 10 – 100
c) less than 10
d) more than 500

Look at the first 30 titles of the papers in your final set and determine which of the following titles are included. Select Yes or No.

Having a hard time finding the correct answer? Watch the video below.

https://youtu.be/Te1pt1jNFFo

1) “A systematic approach to project related concepts of scrum”
2) “Why and how is Scrum being adapted in practice: A systematic review.”
3) “Scrum: The art of doing twice the work in half the time”
4) “Agile project management with Scrum.”
5) “The elements of scrum: A multiple case study”

Narrowing your search results

Narrow your search results by adding PICOC terms

When you search for meta-analysis and/or systematic reviews, one PICOC term (with or without its alternative and related terms) often may suffice to yield relevant studies. The same may be true for controlled and/or longitudinal studies. When your intervention or topic of interest concerns a construct that is well defined (e.g., 360-degree feedback or virtual team), then your search often yields a manageable number of meta-analyses or high-quality studies. In those cases, it does not make sense to limit the results by adding another PICOC term, as you will most likely end up with too few (or even zero) studies.

In some cases, however, the number of results is very large, and reading all the titles and abstracts to see whether the study is relevant is just too time consuming to be practical. For example, when we searched for controlled and/or longitudinal studies on virtual teams, we found more than 80 results. When we search for low-quality studies, the number of results is often much higher, sometimes more than 1,000. In those cases, we should add a second PICOC term (usually the O [outcome] or the P [population]). By adding a second term, we not only limit the number of results but also increase the relevance of the studies found.

Example

Imagine we are interested in whether introducing virtual teams will (positively or negatively) affect the performance of an insurance company. We have defined the following PICOC:

P: Insurance sales agents

I: Virtual working/virtual teams

C: Traditional teams/collocated teams

O: Performance (number of insurance policies sold)

C: Large Dutch insurance company that operates in a highly competitive market

Select / deselect all | Search with AND | Search with OR | Delete Searches

Search ID#	Search Terms	Search Options	Actions
S7	S5 AND S6	**Search modes** - Boolean/Phrase	View Results (12)
S6	TI performance	**Search modes** - Boolean/Phrase	View Results (105,842)
S5	S3 AND S4	**Search modes** - Boolean/Phrase	View Results (88)
S4	TI(experiment* OR controlled OR longitudinal OR randomized OR quasi) OR AB(experiment* OR "controlled stud*" OR "controlled trial" OR "control group" OR "control variable" OR "comparison group" OR "comparative stud*" OR quasi OR longitudinal OR randomized OR randomly OR laboratory OR "before and after stud*" OR "pretest post*" OR "time series" OR "case control" OR "case cohort" OR "cohort stud*" OR "prospective stud*")	**Search modes** - Boolean/Phrase	View Results (363,609)
S3	S1 OR S2	**Search modes** - Boolean/Phrase	View Results (683)
S2	TI "distributed teams" OR AB "distributed teams"	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (111)
S1	TI "virtual teams" OR AB "virtual teams"	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (585)

A search for high-quality studies with the term virtual teams and the related term distributed teams yields more than 80 results. We therefore add a second PICOC term to the search: performance. As you can see in the screenshot below, when we combine the outcome with AND, we get 12 results.

Example

A hospital administrator wants to find out what is known in the research literature about the effect of 360-degree feedback on the performance of physicians. She has defined the following PICOC:

P: Physicians

I: 360-degree feedback

C: No feedback

O: Performance

C: Large university hospital in the United States

*She conducts a search with the terms "***360-degree feedback***" OR "***multisource feedback***" in the title and the abstract. When she filters the combined outcome for meta-analyses and/or systematic reviews, she finds 4 studies, of which only 1 seems relevant. A search for controlled and/or longitudinal studies yields 16 results, of which 4 seem relevant. None of the studies, however, concern physicians or hospitals. She therefore decides to add a second PICOC term: "***hospital****". Because this term may be too narrow, she decides to include the related terms "***healthcare***" and "***clinic****". When she runs the searches and combines the outcome with AND, she gets 6 results, of which 2 seem relevant.*

Select / deselect all | Search with AND | Search with OR | Delete Searches

Search ID#	Search Terms	Search Options	Actions
S12	S3 AND S11	**Search modes** - Boolean/Phrase	View Results (6)
S11	S8 OR S9 OR S10	**Search modes** - Boolean/Phrase	View Results (93,221)
S10	TI clinic* OR AB clinic*	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (29,921)
S9	TI hospital* OR AB hospital*	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (31,553)
S8	TI healthcare OR TI "health care" OR AB healthcare OR AB "health care"	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (46,468)
S7	S3 AND S6	**Search modes** - Boolean/Phrase	View Results (16)
S6	TI(experiment* OR controlled OR longitudinal OR randomized OR quasi) OR AB(experiment* OR "controlled stud*" OR "controlled trial" OR "control group" OR "control variable" OR "comparison group" OR "comparative stud*" OR quasi OR longitudinal OR randomized OR randomly OR laboratory OR "before and after stud*" OR "pretest post*" OR "time series" OR "case control" OR "case cohort" OR "cohort stud*" OR "prospective stud*")	**Search modes** - Boolean/Phrase	View Results (363,609)
S5	S3 AND S4	**Search modes** - Boolean/Phrase	View Results (4)
S4	TI(meta-analy*) OR AB(meta-analy*) OR TI("systematic review") OR AB("systematic review")	**Search modes** - Boolean/Phrase	View Results (7,429)
S3	S1 OR S2	**Search modes** - Boolean/Phrase	View Results (230)
S2	TI "multisource feedback" OR AB "multisource feedback"	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (71)
S1	TI "360 degree feedback" OR AB "360 degree feedback"	**Limiters** - Scholarly (Peer Reviewed) Journals **Search modes** - Boolean/Phrase	View Results (166)

Learn by doing 5.14

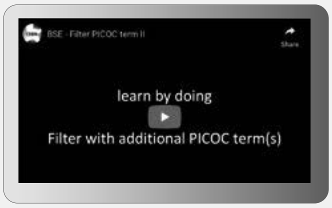
Go to Business Source Elite.

Conduct a search in Business Source Elite for studies with the term *360 degree feedback* **OR** *multisource feedback* **in the title and the abstract. Your search should yield about 130 studies. Now assume that you are particularly interested in whether 360-degree feedback improves the performance of managers and leaders in the public sector. Limit the number of studies by searching for studies with** *managers* **OR** *leaders* **in the title.**

How many peer-reviewed articles does your search yield?

a) About 1 to 5 results
b) About 5 to 10 results
c) About 10 to 50 results
d) About 50 to 100

Having a hard time finding the cor ect answer? Watch the video below.



https://youtu.be/Bv_SJqjYVlE

Did I get this 5.19

The head of a large police organization considers introducing annual performance reviews for all police officers. He asks you to make a summary of what is known in the scientific literature on the effect of performance reviews on people's performance. When you search the Internet and Google Scholar, you find that *performance appraisal** **and** *performance evaluation** **are alternative terms. When you conduct a search in Business Source Elite, you find several meta-analyses and a large number of high-quality single studies. However, none of them are about police officers. You, therefore, decide to do an additional search to find studies on** *performance review****,** *performance appraisal****, OR** *performance evaluation** **in the title or abstract AND that use the term** *police* **in the title or abstract.**

How many peer-reviewed articles does your search yield?

a) About 1 to 10 results
b) About 10 to 25 results
c) About 25 to 75 results
d) More than 75 results

Having a hard time finding the correct answer? Watch the video below.

https://youtu.be/-tET5O7lw2Y

Narrow your search results by limiting the date range

Another way to limit the number of results is by limiting the date range. You can limit the range by adjusting the date slider. After executing your search, the date slider feature is under “Limit To” to the left of the result list. To set a start date for your results, drag the left slider bar toward the middle, and the results list is refreshed. To set an end date, drag the right slider bar toward the middle, and the results list is refreshed. You can return to your original date range by clicking the x icon in the Current Search box to remove the date range limiter.

Research Databases: Some final tips & tricks

- **Keep it simple!**

Often a search with only the academic construct (or your most important PICOC term) AND "meta-analysis" yields more relevant studies than long, detailed search-strings.

- **If you find limited results, search with a broader term.**

For example, if you conduct a search for collaboration and multidisciplinary teams, try also the broader term teams.

- **Split up word combinations**

Instead of searching with the term “performance feedback” (between quotation marks), search for performance AND feedback. In ABI/INFORM Global, the first option yields 125 results, whereas the second option yields 351 results. Always try to imagine how authors may have used the search terms in the title or abstract. For example, if you search only for the term cultural diversity in the title, you would miss the meta-analysis “Unraveling the effects of cultural and gender diversity in teams.”

- **Don’t use too many search terms.**

In the example above, a search for articles with the terms cultur* OR diversity AND performance in the title would also leave out the meta-analysis, because performance is not mentioned in the title. Keep in mind that, in general, meta-analyses and systematic reviews do not mention all outcome measures in the title or abstract, so when your search yields no (or limited) results, consider leaving out a search term.

- **When there are no meta-analyses or systematic reviews, try to search for articles with the word review in the title.**

This will usually yield several review studies that don’t meet the quality criteria of true systematic reviews or meta-analyses, but they can nevertheless be useful.

- **Do not panic when your search yields a large number of studies.**

Skimming through, say, 80 to 100 titles or abstracts can be done pretty quickly. In addition, the chances are that most of the studies will not be relevant to your question (or will be completely incomprehensible), so your final selection will most likely be much smaller.

- **Searching for relevant empirical studies is an iterative process.**

Although this module presents the search process as highly systematical and linear, in practice you will most likely jump back and forth between the steps, especially when your initial search yields unsatisfactory results.

Documenting Your Search Process

As explained in the introduction of this module, an evidence-based search is systematic, transparent, and verifiable so that other people can check or reproduce your search. For this reason, you should clearly document the your search strategy (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.

Documenting your search in a research database

When conducting a systematic and rigorous search in research databases such as ABI/INFORM, Business Source Elite, or PsycINFO, you can document your search by taking a screenshot of your search history (as shown in the examples provided in this module) or by drafting a table. Below is an example of a table that shows how a search for empirical studies on factors influencing the effectiveness of meetings was conducted across three research databases.

ABI/Inform Global, Business Source Elite, PsycINFO peer reviewed, scholarly journals, August 2022			
Search terms	**ABI**	**BSP**	**PSY**
S1: TI(effective*) AND TI(meeting*) OR AB("effective meeting*")	103	86	60
S2: TI(work*) AND TI(meeting*)	107	109	170
S3: TI(lead*) AND TI(meeting*)	25	66	46
S4: TI("group meeting*") OR TI("business meeting*")	91	129	173
S5: TI(meeting*) AND TI(office)	11	17	7
S6: TI(meeting*) AND TI(skills)	16	12	8
S6: S1 OR ... S5, limit > 2000	225	252	151
S7: S6 AND filter meta-analyses	1	3	0
S8: S6 AND filter controlled/longitudinal studies	12	20	6
S9: S6 AND filter quantitative studies	180	176	95
Study selection			
Total number of studies (S7 OR S8 OR S9)	483		
Title and abstracts screened for relevance; duplicates removed	287		
Total number of studies selected	**32**		

Documenting your search in Google Scholar

When conducting a quick and pragmatic search in Google Scholar, it is often not necessary to document your search. However, if you plan to share your findings with others—such as colleagues, or your manager—or if you are conducting a Critically Appraised Topic (CAT), it is important to be able to demonstrate exactly how you carried out your search.

Unfortunately, Google Scholar does not offer built-in features like a search history or an overview of previous searches. Therefore, to document your search systematically and transparently, you should include the following:

1. **Date of Search:** Google Scholar's content changes frequently, so always note the date you conducted the search.
2. **Search Terms and Queries:** Record the exact search terms and queries you used.
3. **Search Settings:** Mention any filters applied, such as date range (e.g., since 2019) or language restrictions (if applicable).
4. **Number of Results Screened:** Indicate how many results you reviewed per query (e.g., the first 50 hits).
5. **Studies Selected:** List the studies you identified as most relevant, preferably in APA style. You can do this by clicking 'Cite' under each title and copying the reference.

Google Scholar

1. **Date of Search:** April 16, 2024
2. **Search Terms and Queries**

   Search 1: effective meetings meta-analysis
   Search 2: effective meetings factors
   Search 3: group meetings effectiveness
   Search 4: effective leading meetings
   Search 5: strategies effective meetings

3. **Search Settings**: Time range: 2000 – 2024
4. **Number of Results Screened:** 250 (first 50 results from each search query)
5. **Studies Selected**: 4

   Allen, J. A., Tong, J., & Landowski, N. (2021-2). Meeting effectiveness and task performance: meeting size matters. [Meeting effectiveness and performance]. The Journal of Management Development, 40(5), 339- 351

   Baran, B. E., Shanock, L. R., Rogelberg, S. G., & Scott, C. W. (2012). Leading Group Meetings: Supervisors' Actions, Employee Behaviors, and Upward Perceptions. Small Group Research, 43(3), 330-355.

   Leach, D. J., Rogelberg, S. G., Warr, P. B., & Burnfield, J. L. (2009). Perceived Meeting Effectiveness: The Role of Design Characteristics. *Journal of Business and Psychology, 24*(1), 65-76.

   Rogelberg, S. G., Leach, D. J., Warr, P. B., & Burnfield, J. L. (2006). "Not another meeting!" Are meeting time demands related to employee well-being? *Journal of Applied Psychology*, *91*(1), 83.

Documenting your search in AI search tools

Most AI tools such as Consensus automatically save your past queries and search results, allowing you to revisit and document previous research findings. In Consensus you can access your search history by navigating to the left sidebar and clicking the > button at the bottom.

To systematically and transparently document your search, include the following:

1. Date of Search: Note the date you conducted the search.
2. Search Queries: Record the exact question(s) or phrases you used.
3. Search Settings: Mention any filters applied, such as date range (e.g., published since 2010) or research method (e.g., meta-analyses).
4. Number of Results Screened: Indicate how many studies you reviewed per query (e.g., the first 20 studies).
5. Studies Selected: List the studies you identified as most relevant, preferably in APA style. To do this, click the ‘Cite paper’ icon located below the title on the right side, and copy the reference.

Consensus

1. **Date of Search:** April 16, 2024

2. **Search Queries**

   Question 1: What are factors that influence the effectiveness of meetings?
   Question 2: What are key design characteristics influencing meeting outcomes?
   Question 3: What role does leadership play in the effectiveness of team meetings?

3. **Search Settings:** Published since 2000, meta-analyses, observational studies

4. **Number of Results Screened:** 30 (first 10 results from each search query)

5. **Studies Selected**

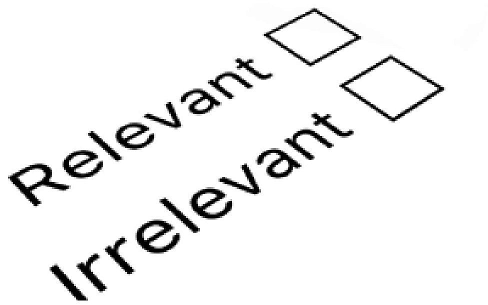
   Hinkin, T., & Tracey, J. (2003). The Service Imperative: Factors Driving Meeting Effectiveness. Cornell Hospitality Quarterly, 44, 17 - 26.

   Nixon, C., & Littlepage, G. (2005). Impact of meeting procedures on meeting effectiveness. Journal of Business and Psychology, 6, 361-369.

   Leach, D., Rogelberg, S., Warr, P., & Burnfield, J. (2009). Perceived Meeting Effectiveness: The Role of Design Characteristics. Journal of Business and Psychology, 24, 65-76.

   Bang, H., Fuglesang, S., Ovesen, M., & Eilertsen, D. (2010). Effectiveness in top management group meetings: the role of goal clarity, focused communication, and learning behavior. Scandinavian journal of psychology, 51 3, 253-61 .

Screening the articles found for relevance



Whether you do a quick search in Google Scholar or Consensus, or conduct a systematic and thorough search in a research database, your search will often yield many studies, some of which will not be relevant to your question and PICOC. The final step is hence to screen the articles to check whether they are relevant.

Screening for relevance is usually a two-stage process. First, compare each title and abstract against your question and PICOC. Unfortunately, not all abstracts will contain the information you need to determine whether the article is relevant. In that case, you need to retrieve the full text and skim through it.

As mentioned earlier, your PICOC will help you to determine whether the findings of a study will be generalizable and applicable to your professional context. Keep in mind, though, that what works in one narrowly defined setting sometimes might not work in another, but that many principles are generalizable to most people or situations. For example, what if you would like to know whether 360-degree feedback will be effective as a tool for improving the task performance of physicians in a Dutch university hospital, and the outcome of your search yields only high-quality studies in which the effect was examined on the performance of American lawyers and German teachers? Would you consider the outcome of these studies, given your question and PICOC, to be relevant? Unfortunately, there are no general guidelines to help you to evaluate the generalizability of research findings, so this is where your professional judgment comes in.

Ten steps in the search process

Step 1: Determine the one or two most important terms in your PICOC.

Step 2: Identify alternative/related terms.

Step 3: Determine whether there are corresponding academic terms.

Step 4: Pretest your search terms and determine which terms yield the most relevant articles.

Step 5: Determine whether a quick search in Google Scholar or Consensus will suffice, or if an extensive search in a research database is required.

Step 6: Search for meta-analyses and/or systematic reviews

Step 7: Search for high-quality studies

Step 8: Search for low-quality studies

Step 9: Limit the number of search results by adding a PICOC term.

Step 10: Screen the titles and the abstracts of the articles found for relevance.

Podcast: ACQUIRE Evidence from the Scientific Literature

In this podcast host Karen Plum discusses with Eric Barends, Managing Director of CEBMa, Denise Rousseau, Professor at Carnegie Mellon University, and Barbara Janssen, Board Member and Fellow of CEBMa how to gather evidence from the scientific literature that will help us address the question we are trying to answer.




In this podcast we discuss the process of searching for academic studies; the difference between everyday /managerial terms and academic constructs; how to find the right terms to search for; the importance of obtaining studies from the right sources, most especially peer reviewed academic journals; the difference between journals and “magazines”; and the importance of documenting the process you followed to identify your evidence base.

We also explore how the peer review process and the meta-analysis help us get to the best available evidence.

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

References

1. Hanratty, J. (February 12, 2018). *What is the difference between a systematic review and a meta-analysis?* [blog post]. CampbellCollaboration. http://meta-evidence.co.uk/difference-systematic-review-meta-analysis.

Learn by doing & Did I get this? Answers and Feedback

Did I get this? 5.1

1: True > Research articles submitted to peer-reviewed journals are first evaluated and critiqued by independent, anonymous scientists in the same field (peers) to determine whether they merit publication in a scientific jou nal.

2: False > This is a common misconception. Unfortunately, poor-quality studies suffering from methodological flaws, bias, and incomplete conclusions a e sometimes rife in peer-reviewed journals.

3: False > This is a common misconception. Sometimes, a research paper is seriously flawed, even when it was conducted by a top university professor and published in a top academic journal.

4: Not quite right > Invalid and unreliable research articles can also be found in peer-reviewed journals, but the process of peer review gives you some assurance that the article is not seriously flawed

Learn by doing 5.1

P: a > The population of interest concerns factory workers.

I: d > The intervention/method of interest concerns on-site chair massage therapy.

C: c > In this case, the intervention (chair massage therapy) is compared to business as usual (continue working).

O: b > The outcome of interest is reduced stress and anxiety.

C: d > In this case, the organization concerns a large midwestern US manufacturing organization that has experienced multiple restructurings and downsizings.

Did I get this? 5.2.

P: a > In this case, the population of interest are software developers.

I: d > In this case the intervention of interest concerns open office design

C: d > The intervention (open office design) is compa ed to individual workspaces with half walls.

O: a > The outcome of interest is whether employees get distracted.

C: b > The context is a small Italian software company.

Learn by doing 5.2

1. I > First, we want to know whether coaching has a positive effect on people's individual performance. If so, then we want to know whether the same holds true clinical managers in an academic hospital specifically.
2. O > First, we want to know whether coaching has a positive effect on people's individual performance. If so, then we want to know whether coaching also improves the performance of clinical managers in an academic hospital.

Learn by doing 5.3

1. P > In this case, the population (knowledge workers) is one of the two most important terms.
2. O > First, we want to know which factors have a positive effect on the performance of knowledge workers in general.

Did I get this? 5.3

1. I > First, we want to know whether knowledge sharing has a positive effect on the innovativeness of teams. If so, then we want to know whether the same holds true for design teams specifically.
2. O > First, we want to know whether knowledge sharing has a positive effect on the innovativeness of teams. If so, then we want to know whether the same holds true for design teams specifically.

Did I get this? 5.4

1. 'Multi-source Feedback' is a widely used academic term.
2. The term 'Full-Circle Feedback' is less common and only occasionally used as a synonym.
3. 'Multi-rater Feedback' is a widely used academic term.
4. The term 'Panoramic Feedback' is not an academic term.

Learn by doing 5.4

The term 'change management' is used to refer to a wide range of activities—from implementing new technology to managing mergers or large-scale organizational culture change—often without a precise definition. In addition, there is considerable overlap with other concepts such as organizational development, strategic transformation, or even project management, making the boundaries unclear. Finally, the term is used inconsistently in practice: different fields (e.g., consulting, HR, academia) may interpret and apply it in different ways. So, while change management is a widely used term, its conceptual clarity is relatively low.

Did I get this? 5.5

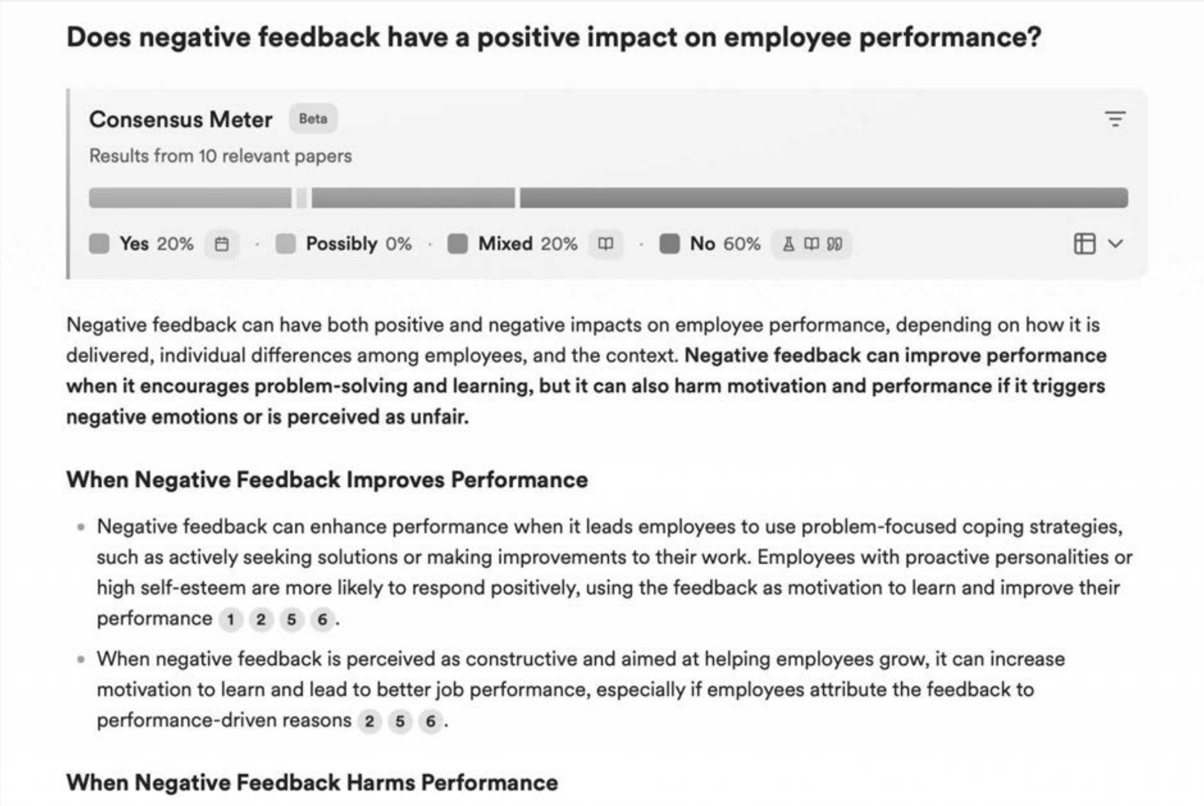
The term 'span of control' is clearly defined as the number of employees who directly report to a manager or supervisor, and it can be reliably measured by simply counting these direct reports. In addition, the term is used fairly consistently across academic, management, and HR literature. As such, its conceptual clarity is relatively high.

Learn by doing 5.5

Most abstracts clearly state that job insecurity negatively affects employee performance (as well as motivation, organizational commitment, and well-being).

Did I get this? 5.6.

As you can see in the screenshot of the outcome of our search in Consensus, negative feedback can have both positive and negative effects on employee performance.



Does negative feedback have a positive impact on employee performance?

Consensus Meter Beta

Results from 10 relevant papers

Yes 20% · **Possibly** 0% · **Mixed** 20% · **No** 60%

Negative feedback can have both positive and negative impacts on employee performance, depending on how it is delivered, individual differences among employees, and the context. **Negative feedback can improve performance when it encourages problem-solving and learning, but it can also harm motivation and performance if it triggers negative emotions or is perceived as unfair.**

When Negative Feedback Improves Performance

- Negative feedback can enhance performance when it leads employees to use problem-focused coping strategies, such as actively seeking solutions or making improvements to their work. Employees with proactive personalities or high self-esteem are more likely to respond positively, using the feedback as motivation to learn and improve their performance 1 2 5 6.
- When negative feedback is perceived as constructive and aimed at helping employees grow, it can increase motivation to learn and lead to better job performance, especially if employees attribute the feedback to performance-driven reasons 2 5 6.

When Negative Feedback Harms Performance

Did I get this? 5.7

A. A systematic and rigorous search in multiple research databases

Whether or not to implement autonomous teams is an important decision that involves significant resource investment and will affect a large number of people. For this reason, a systematic and rigorous search in multiple research databases is required.

Did I get this? 5.8

B. systematic and rigorous search in multiple research databases

In this scenario, you simply want to check whether there is relevant research that supports your colleague’s claim—there is no major decision to be made yet that involves significant resources or affects a large number of people.

Learn by doing 5.6

1a: Incorrect > Document 3 contains the terms *red* and *blue*, but so does another document.

1b: Correct > Both document 3 and document 6 contain the terms *red* and *blue*.

1c: Incorrect > Documents 3, 5, 6, and 7 all contain the term *red*, but only documents 3 and 6 contain both *red* and *blue*.

1d: Incorrect > Documents 2, 3, 4, 5, 6, and 7 contain the term red or blue, but only documents 3 and 6 contain both *red* and *blue*.

2a: Incorrect > Document 1 contains the terms *pink* and *purple*, but documents 1, 2, 4, 5, and 8 contain pink or *purple*.

2b: Incorrect > Both document 1 and 4 contain the term *pink*, but documents 1, 2, 4, 5, and 8 contain *pink* or *purple*.

2c: Incorrect > Documents 1, 2, and 5 all contain the term *purple*, but documents 1, 2, 4, 5, and 8 contain either *pink* or *purple*.

2d: Correct > Documents 1, 2, 4, 5, and 8 contain the term *pink* or *purple*.

Did I get this? 5.9

a: Incorrect > Document 3 contains the terms *brown* and *green*, but so do documents 7 and 8.

b: Incorrect > Document 5 does not contain the term *brown* or *green*.

c: Correct > Documents 3, 7, and 8 contain the terms *brown* and *green*.

d: Incorrect > Documents 1, 2, 3, 4, 6, 7, and 8 contain the term *brown* or *green*, but only documents 3, 7, and 8 contain the both *brown* and *green*.

Did I get this? 5.10

a: Correct > Documents 2 and 4 contain the terms green and a variation of the term *perform**.

b: Incorrect > Documents 3 and 6 do not contain the term green or a variation the term *perform**.

c: Incorrect > Documents 2 and 4 contain the terms *green* and *perform**, but document 7 contains the term *perforation*, which is not a variation of *perform**.

d: Incorrect > Documents 1, 2, 3, 4, 5, 7, and 8 contain the term *green* or a variation of the term *perform**

Did I get this? 5.11

a: Incorrect > It is not likely that the database made a calculation error, so there must be another reason.

b: Correct > For example, some authors may have explained in the abstract that 360-degree feedback is also known as multisource feedback. In the individual searches, this article is counted as two separate articles and, in the combined search, as one.

c: Incorrect > If only articles that use more than one search term in the title or the abstract are included in the outcome of the combined search, then the total number of results would be much lower than 483.

d: Incorrect > If articles that use more than one search term in the title or the abstract were excluded, then the total number of results would be lower than 483.

Learn by doing 5.7

a,b,c: Incorrect > Check the video to see where you made a mistake.

d: Correct > Your search should yield more than 4,000 results. Watch the video to see how we searched.

Did I get this? 5.12

a, c,d: Incorrect > Check the video to see where you made a mistake.

b: Correct > Your search should yield between 400 and 1000 results. Watch the video to see how we searched.

Did I get this? 5.13

a: Incorrect > If all articles that use more than one search term in the title or the abstract were included in the outcome of the combined searches, then the total number of results would be much lower than 3,779.

b: Incorrect > You are right that adding more PICOC terms is a way to limit the number of results, but that is not the reason the combined search yielded so many results.

c: Correct > Because OR was used, the combined search yielded all papers with the term physicians, multisource feedback, 360-degree feedback, or multirater feedback in the title or the abstract.

d: Incorrect > If you click the Search button twice, you get the same results twice.

Learn by doing 5.8

a: Yes > A search with *entrepreneur** in the title yields more than 19,000 results, most of which seem to be relevant.

b: No > A search with *enterpriser** in the title yields only few results, most of which seem to be irrelevant.

c: No > A search with *industrialist** in the title yields limited results, most of which seem to be irrelevant.

d: Yes > A search with "*business owner**" in the title yields more than 350 results, most of which seem to be relevant.

e: No > A search with *tycoon* in the title yields only a few results, most of which seem to be irrelevant.

f: Yes > A search with "*founder CEO**" in the title yields only a few results, but most seem to be relevant.

Did I get this? 5.14

a: No > A search with "*financial reward**" in the title yields only 20 to 40 results, most of which seem to be irrelevant.

b: No > A search with "*financial fee**" yields in the title yields only a small number of (irrelevant results.

c: Yes > A search with "*financial incentive**" in the title yields more than 300 results, most of which seem to be relevant

d: Yes > A search with "*pay for performance plans*" in the title yields only a few results, but most seem to be relevant

e: No > A search with "*performance reward**" in the title yields only a few results, most of which seem to be irrelevant

f: No > A search with "*variable pay*" in the title yields only 40 to 60 results, and most seem to be irrelevant

Did I get this? 5.15

Our answer: The person searched with performance OR feedback in the title (S7. The Boolean operator AND should have been used instead of OR.

Learn by doing 5.9

a: Yes > This paper should be in your final set, because it uses both *multisource feedback* and *meta-analysis* in the title.

b: No > This paper should not be in your final set, because it does not use *meta-analysis* or *systematic review* in the title or abstract.

c: No > This paper should not be in your final set, because it does not use *meta-analysis* or *systematic review* in the title or abstract.

d: Yes > This paper should be in your final set, because it uses *multirater feedback* in the title and *meta-analytic* in the abstract.

Learn by doing 5.10

a: Yes > Your search should yield about 3 to 10 results. Watch the video to see how we have searched. b,c,d: No > Check the video to see where you made a mistake

Learn by doing 5.11

a: Yes > This paper should be in your final set

b: No > This paper should not be in your final set

c: No > This paper should not be in your final set

d: Yes > This paper should be in your final set

e: No > This paper should not be in your final set

Did I get this? 5.16

a: Yes > This paper should be in your final set

b: No > This paper should not be in your final set

c: No > This paper should not be in your final set

d: Yes > This paper should be in your final set

Learn by doing 5.12

a: Yes > This paper should be in your final set

b: Yes > This paper should be in your final set. Note that this paper is a meta-analysis but nevertheless was identified by the filter for controlled and/or longitudinal studies because the term *longitudinal studies* is used in the abstract.

c: No > This paper should not be in your final set, because the title or the abstract does not use any of the terms that are included in the search filte .

d: Yes > This paper should be in your final set

Did I get this? 5.17

a: No > This paper should not be in your final set

b: Yes > This paper should be in your final set

c: No > This paper should not be in your final set

d: Yes > This paper should be in your final set

Did I get this? 5.18

a,c,d: Incorrect > Check the video to see where you made a mistake.

b: Correct.

1: Yes > This paper should be in your final set

2: Yes > This paper should be in your final set

3: No > This paper should not be in your final set

4: Yes > This paper should be in your final set

5: No > This paper should not be in your final set

Learn by doing 5.13

a,b,d: Incorrect > Check the video to see where you made a mistake, and try again.

c: Correct.

Did I get this? 5.19

a,b,d: Incorrect > Check the video to see where you made a mistake, and try again.

c: Correct.

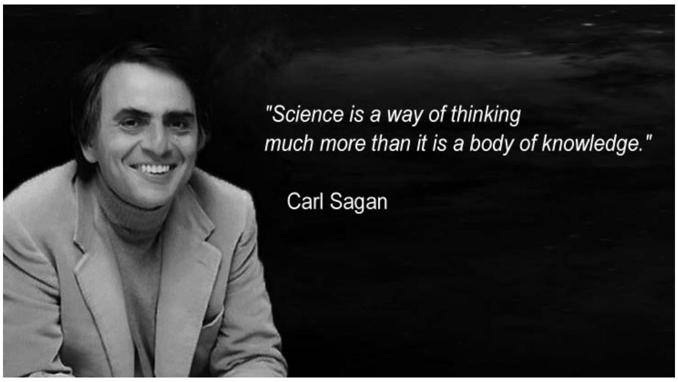
MODULE 6

A Short Introduction to Science

Learning objectives:

- Determine whether a study meets the standards of good science.
- Determine whether a study's findings are practically relevant.
- Determine whether bias, confounding, or effect modification may have affected the results.
- Determine a study's research design.
- Efficiently read a research article.

What is science?



What comes to mind when you hear the term ***science***? Laboratory workers in white coats? The Large Hadron Collider in Geneva? Memories of exciting and seemingly dangerous chemistry experiments at school? Or mice in a cage? These are widespread and popular ideas of science. Is it any surprise, then, that many managers and leaders, when confronted with the term ***scientific*** research, wonder what it has to do with them or their work? Of course, science isn't only about what might be called ***hard*** science. As a method and way of thinking about how to understand the world, science can be applied to almost anything, including organizations, management, workers, and business.

The basic purpose of science is to acquire information that will help us to describe, explain, predict, and control phenomena in the world. According to astrophysicist Niel deGrass Tyson, "Science distinguishes itself from other human pursuits by its power to examine and understand phenomena on a level that allows us to predict with varying degrees of accuracy, if not sometimes control, the outcomes of events in the natural and human-made world." (1) For science to do so, however, we need trustworthy information or data, acquired in ways that minimize bias and other misleading factors.

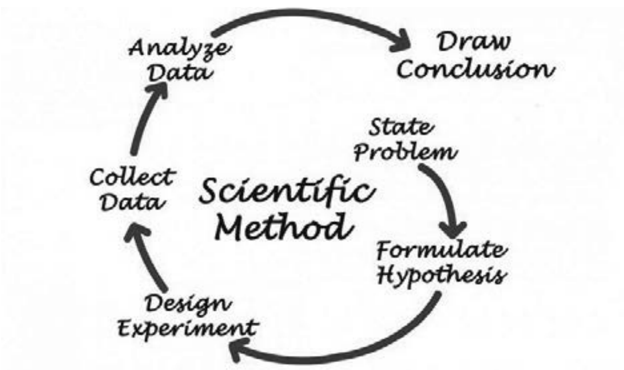
Nevertheless, as human beings, we are inclined to use the most easily accessible source of information: ourselves. We rely on what we remember from our own experiences, what we think, and what we believe to be true. Our self-centered approach to information saves time and effort – and probably works reasonably well for simple day-to-day decisions. For more complicated decisions in business and management, however, relying solely on our own experience and judgment may well lead to poor decisions. As you learned in Module 4, APPRAISE Evidence from Practitioners, we are prone to cognitive biases in our thinking, and this causes us to make mistakes in analyzing and interpretating our own experiences and judgments. Science emerged in part as a response to the twin problems of relying solely on personal information and the biases inherent in interpreting it. A core activity of science, therefore, is gathering objective, external information rather than relying solely on the subjective internal knowledge in our heads. Even the smartest people can easily be fooled into believing something that is not true, so we need to put some safety checks into place when we acquire and appraise external information.

Three safety checks

1. The scientific method

An important safety check is the *scientific method*, a defining feature of science since the 17th century. The scientific method can be summarized in one sentence: *Do whatever it takes to avoid fooling yourself into thinking something is true when it is not or that something is not true when it is.*(1) Scientists use this safety check to ensure the strength/trustworthiness of their findings usually by following these steps

1. Ask a question about something they observe (how? what? when? how many? who? why? or where?).
2. Formulate a hypothesis (an assumption about how things work or a prediction about what will happen).
3. Test the hypothesis by doing an experiment or making systematic observations.
4. Collect the data.
5. Analyze the data.
6. Draw a conclusion as to whether and the extent to which the hypothesis is likely to be right.



Did I get this 6.1

Read the scenario below.

Employee turnover can have a major impact on an organization: when employees leave, the organization suffers many costs. For this reason, a scientist decides to do research on the major reasons for employees leaving an organization.

Order the scientist's activities according to the steps of the scientific method.

A. Interview HR managers at several organizations. Ask them what they think are major reasons that employees leave. Most managers feel that employees who are bored with their job are the most likely to leave the organization.

B. Collect the answers provided to the survey and the organizations' average employee turnover rate.

C. Formulate a hypothesis. Employees who feel they have a monotonous job, with seemingly unimportant tasks, will most likely leave the organization. Therefore, organizations with many employees perceiving their job as boring will have a high employee turnover rate.

D. Draw a conclusion as to whether organizations with many employees perceiving their job as boring have a high employee turnover rate.
E. Calculate a correlation between perceived boredom and turnover rate, running a regression to see whether (and to what extent) boredom predicts employee turnover.
F. Distribute a survey that measures whether people perceive their jobs as boring among a large sample of employees from 20 organizations.

2. Organized skepticism (peer review)

Another safety measure that lies at the core of science is that evidence generated by scientists is subject to *organized skepticism*. Findings are scrutinized collectively – by the scientific community – f om a position of distrust. The burden of proof is on the scientist making a novel claim. In this sense, science is intrinsically cautious. As Naomi Oreskes, in her TED talk "Why We Should Trust Scientists" states, "It is quite hard to persuade the scientific community to say, 'Yes, we know something, this is true.'" (2)

So, another way to think of science is to see it as the consensus of scientific experts, who – through a process of collective scrutiny – have judged the evidence and come to a conclusion about it. Organized skepticism does not mean unanimity, in that we expect scientists to continue to question and raise alternative explanations in pursuit of deeper understanding. However, consensus means that a general agreement exists without strong arguments to the contrary. Note, however, the organized skepticism of science means that a theory is never proven, as disconfirming evidence is always considered a possibility. However, confidence can be high and uncertainty low where scientific consensus is strong.

3. Replication

The final safety check to ensure the strength/trustworthiness of scientific claims is replication. Repeating studies to see if the same result is obtained using exactly the same method is a cornerstone of science. If novel findings from scientific research can be replicated, it means they are more likely to be correct. Multiple replications of scientific findings may turn a hypothesis into a more formal statement or theory. On the other hand, if the findings cannot be replicated, they are likely to be incorrect or oversimplified (due to some error or even chance). The following example shows why replication is essential in science. (3)

Example

In 1998, a British researcher published an article in a medical journal reporting that he had found a link between a common childhood vaccine and autism. According to the article, children in his study developed autism soon after receiving the vaccine. (4)

Following publication of the article, many parents refused to have their children vaccinated. Several epidemics occurred as a result, and some children died. Soon after the original study was published, other researchers failed to replicate its findings: no other studies could find a link between the vaccine and autism. Eventually, researchers found that the original study was a fraud.(5)

The author had received a large amount of money to find evidence that the vaccine caused autism, so he faked his results. If other scientists had not tried to replicate the research, the truth might never have come out.

Did I get this 6.2

Read the scenario below.

A scientist wants to know the predictors of successful leadership in the technology industry. She therefore examines the characteristics of eight successful leaders: Steve Jobs (Apple), Michael Dell (Dell), Bill Gates (Microsoft), Mark Zuckerberg (Facebook), Larry Ellison (Oracle), Jan Koum (WhatsApp), Travis Kalanick (Uber), and Evan Williams (Twitter).

After examining a large number of variables, such as age, personality traits, and leadership style, she finds a characteristic that all eight leaders have in common: they all dropped out of college or university. Based on this finding, she concludes that higher education does not contribute to successful leadership and publishes her findings in a popular management magazine.

Does this research meet the standards of good science? Check the three most important weaknesses.

a) No, because the scientist did not submit her research for peer review.

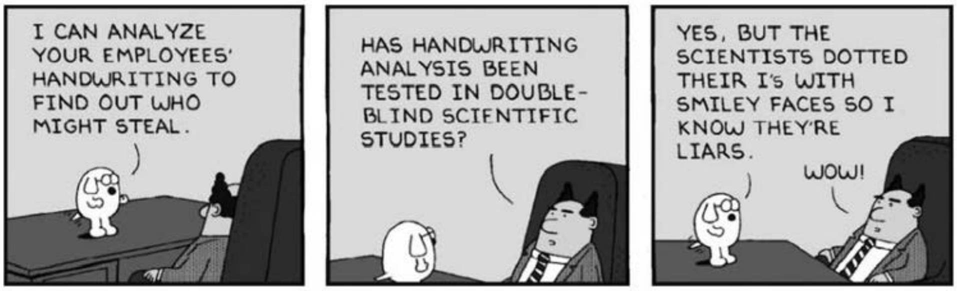
b) No, because the scientist should not have limited her research to the technology industry.

c) No, because the scientist did not try to replicate her findings

d) No, because the scientist should have searched for more common characteristics.

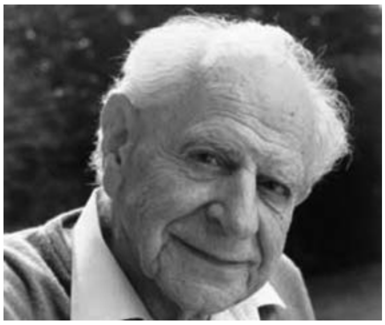
e) No, because the scientist did not follow the scientific method: she formulated a hypothesis after the collection and analysis of the data.

Pseudoscience



Do you know what pseudoscience is? Disciplines such as astrology and parapsychology are regarded as pseudo sciences. You know from our discussion so far that science is a method used to test hypotheses in a way that takes account of coincidence, bias, and other misleading factors. It follows that if we cannot test a hypothesis or theory, we cannot subject it to scientific investigation.

One of the people who has thoroughly examined the difference between science and pseudoscience is the philosopher Karl Popper. He states that it is easy to obtain evidence in favor of virtually any theory. (This seems even more true in the Internet era, where an array of evidence and ideas for almost any claim or theory can be found via a brief online search.) According to Popper, a theory should be considered scientific only if it is the positive result of a genuinely “risky” prediction, which might conceivably have been found to be false. Put differently, a theory or model is scientific only if it can be tested and falsified.



Karl Popper, 1902 – 1994

Take, for example, Uri Geller, a famous psychic who repeatedly demonstrated on television and on stage that he could bend keys and restart watches by using “mental energy.” However, when his assumed psychic abilities were tested scientifically, Geller stopped the experiment by claiming that the scientific setting interfered with his mental energy, making it impossible to refute or falsify his claim.

Did I get this 6.3

Read the scenario below.

A university professor tells you that students learn better when teaching is tailored to their individual "learning styles"—visual, auditory, or kinesthetic. She bases this belief on the large number of publications on the internet and in popular magazines promoting the learning styles theory, as well as on her own classroom experience and student feedback.

Which of the following concerns should raise red flags that learning styles might not be a scientifically supported theory?

a) The theory is promoted primarily in popular media rather than peer-reviewed scientific journals
b) The professor relies on personal experience and student feedback instead of scientific evidence.
c) The categories used by the theory (visual, auditory, kinesthetic) are not clearly defined and hard to measure.
d) All of the above

Science is not about truth or proof

Developing a scientific understanding of the trustworthiness (validity and reliability) of information is important. But equally important is the realization that science is not about truth or proof. Science is about gathering information and testing assumptions (hypotheses) in ways that allow us to estimate how *likely* it is that something is true. We can never know for sure. Our uncertainty stems from three main sources:

1. It is always possible that new information will cast serious doubt on a well-established model or theory. For example, more rigorous research may demonstrate that the underlying assumptions are incorrect or that previous research was flawed in ways that produced biased or even false results. To claim something is true or proven is to miss the point of science: research can deliver only the best current evidence and calculate a probability, but when new evidence becomes available, this probability may change. The job of an evidence-based manager is to make decisions on the basis of best available evidence at a given time while remaining open to new and better evidence in the future.
2. Even when something seems very close to being proven, it is still subject to boundary conditions – it always depends on the situation. Even though a lot of data may support a particular theory, there may be other data from other settings to suggest that this theory does not hold true everywhere. Goal-setting theory, which proposes that setting moderately difficult goals leads to higher performance, for example holds up well in many contexts but less so in others. (6)
3. As we learn more and more about a given model or theory, we sometimes discover that our original findings were not quite correct – or at least were not specific enough. Take the concept of *organizational commitment*: when first developed, it was a general and one-dimensional construct that suggested that employees were either more committed or less committed to their organizations. However, subsequent research has revealed at least three different and specific forms of commitment, which have different and specific effects on outcomes. (7) To claim, therefore, that commitment in general has some effect on outcomes no longer holds water, as it depends on which form of commitment we are looking at.

If you dig into scientific findings looking for absolute truth and proof, then you will be disappointed. If you look for evidence about likelihoods and probabilities, then research findings can be very useful to overcoming the limitations of human judgment. We describe five of these limitations in detail in the next pages.

Limitation 1: Coincidence

The first limitation of human judgment that science aims to overcome is coincidence: Could our observation of a phenomenon be due simply to chance? For example, when we notice that a person becomes ill after eating wild berries, we are inclined to conclude that the person became sick because of eating the berries. However, this could also just be coincidence. To rule out chance, we therefore apply the scientific method: based on our observations, we formulate a hypothesis (in this case, that eating wild berries makes you sick), and then test our hypothesis by doing an experiment. However, as you just have learned, science is not about proof or the truth but about probabilities and likelihoods. Thus, before we run our experiment – or conduct any other type of study – we should first determine what degree of uncertainty we are willing to accept. Should we accept a probability of 10 percent that our experiment's outcome was due to chance, or should we be more lenient and accept a threshold of 20 percent? Or an even higher value?

Statistical significance: p-value

In 1925, the English statistician Ronald Fisher suggested that, within the realm of science, this threshold should be set at 5 percent (one in twenty). (8) This threshold was – unfortunately – later referred to as the ***significance level***, and the corresponding probability (p) as p-value. (9) Fisher argued that if the p-value is higher than .05, then the probability that a study's outcome is due to chance should be considered too high. From that moment onward, the significance level of p = .05 became the most widely used but also most misapplied and misunderstood statistic in science. This is because, from an evidence-based perspective, there is a serious problem with this metric.

Statistical significance versus practical relevance

In the realm of science, a significant outcome is often interpreted as a finding that was most likely not due to chance. (10) In daily life, however, significant means 'sufficiently g eat or important to be worthy of attention'. However, statistically significant research outcomes are sometimes insignificant f om a practical perspective (and vice versa). This is because statistical significance, outcome and sample size a e interlinked. If the effect found is small but the sample size is very large, the p-value can be statistically significant. Similarly, if the effect is large and the sample size is small, the p-value can also be significant. Thus, when you make a sample size large enough, even highly trivial outcomes can be statistically significant. (11)

Example

Imagine that someone has developed a training program that aims to increase the IQ of young children. With only four children enrolled in the program, an increase of 10 IQ points would be a statistically significant outcome. However, with 25 children enrolled, an increase of 4 IQ points would also be significant, and with 10,000 children even an increase of 0.2 IQ points would be significant (see table on next page).

Sample size (total number of children)	**Effect size** (significant increase in IQ)
4	10
25	4
100	2
10,000	0.2

Would you send your child to this training program? Probably not, as an increase of 0.2 IQ points almost certainly has no practical relevance. This means that the fact that a study's outcome is statistically significant has limited meaning and is of limited value, because it doesn't tell us if that outcome is of practical relevance.

From a practical perspective, a significance level of p = .05 is often too strict!

Learn by doing 6.1

Read the following scenario.

A study reports a 10 percent lower rate of production-line errors in manufacturing companies that use a particular quality management model. The p-value for the difference in error rate (compared with companies that have not implemented the model) is 0.07. The common interpretation of this p-value is that there is a probability of 7 percent that this outcome is due to chance. Now, imagine that every 1 percent error decrease yields a profit increase of $50,000, which equates a total of $500,000. Let's also assume that the cost of implementing the model is quite low. Finally, the study's research design is qualified by peer reviewers as "rigorous."

Should you implement the model?

a) No

b) Yes

c) I need more information

Did I get this 6.4

Read the following scenario.

A study reports a 0.12 percent lower absenteeism rate in companies that use a new prevention model. The p-value for the difference in absenteeism rate (compared with companies that do not use the prevention model) is 0.0001, suggesting that there is only a very small probability (0.01 percent) that this finding is due to chance. Now, let's assume that the cost of implementing the model is quite substantial.

Should you implement the model?

a) No

b) Yes

c) I need more information

Limitation 2: Methodological bias

The second limitation of human judgment that science aims to overcome is bias: Could our observation be due to personal preference or prejudice? In Module 4, APPRAISE Evidence from Practitioners, you learned that human judgment is prone to a wide range of cognitive biases, such as the illusion of causality, confirmation bias, availability bias, and outcome bias. Consequently, people (including researchers and scientists) can be easily fooled into believing something that is incorrect. Thus, when researchers study a phenomenon or test a hypothesis, they need to put a safety measure into place: the scientific method. Unfortunately, even the scientific method – or the way science in general is practiced – is p one to all kinds of different methodological biases. Some of the most common are described next.

Selection bias

Also called ***sampling bias***, selection bias occurs when the particular choice of participants in a study leads to an outcome that is different from the outcome that would have occurred if the entire population were studied. For example, if we want to know something about people's attitudes toward sex outside of marriage, surveying people in a nightclub on a party island resort would likely yield a result different from a survey of churchgoers in the US Bible Belt. Researchers use ***probability sampling*** *(or random sampling)* to prevent selection bias. Note that *random* means the people in the sample are chosen by chance (thus, each person in the population has the same probability of being chosen). When you pick a truly random sample, you reduce the chances of selection bias.

Social desirability bias

This bias occurs when research participants answer questions in ways that they think are most socially acceptable. Social desirability reflects the concern people may have about how others view them. It is a difficult bias to overcome because people are inclined to report inaccurately on sensitive or personal topics in order to present themselves in accordance with other people's expectations. It can be reduced in several ways, such as by assuring confidentiality or anonymity, by observing behavior directly, or by using qualitative methods that build trust between researcher and study participants.

Halo effect

The halo effect describes the basic human tendency to make generalized inferences based on a few pieces of information. In his article "The Halo Effect and Other Business Delusions," Phil Rosenzweig gives the following example: "When a company is doing well, with rising sales, high profits, and a surging stock price, observers naturally infer that it has a smart strategy, a visionary leader, motivated employees, excellent customer orientation, a vibrant culture, and so on. When that same company suffers a decline – when sales fall and profits shrink – many people a e quick to conclude that the company's strategy went wrong, its people became complacent, it neglected its customers, its culture became stodgy, and more." (12) This bias is also often present in employee performance ratings and employee selection. In good scientific research, we can minimize the halo effect by blinding (or hiding) participants' characteristics, such as name, age, and gender in the case of individuals or reputation, brand, and profitability in the case of organizations

Learn by doing 6.2

Read the excerpt from the following abstract from a research article.

We identified eight companies that had a sustained period of five years over which the cumulative total stock return dramatically outperformed the general market and its competitors. We then identified eight companies that had a sustained period of five years over which the cumulative total stock return dramatically underperformed the general market and its competitors. A group of 5 reviewers independently evaluated the companies' annual reports, company records and annotated interviews with senior managers to assess the companies' Talent Management (TM) practices. It was found that the eight outperformers all deployed a wide range of TM practices, whereas six underperformers did not deploy any TM practices. This result clearly indicates that companies that deploy TM practices perform better financially.

Could there be methodological bias that may have affected the results?

1) Yes, selection bias.
2) Yes, social desirability bias.
3) No, there is no clear indication of bias.
4) Yes, halo effect.

How could the researchers have minimized bias?

a) By blinding (or hiding) the companies' size and annual turnover.
b) By picking a truly random sample
c) By assuring the participants' confidentiality or anonymity
d) By blinding (or hiding) the companies' name and financial performance

Did I get this 6.5

The following statement was part of a survey that was used in a study on nurse leadership.

My team manager is very conscious of risk and shows strong leadership in ensuring we manage risk in the right way. He/she is an excellent role model for me.

Strongly disagree / Disagree / Neither agree nor disagree / Agree / Strongly agree

The questionnaire was distributed among a random sample of 10,000 nurses from 120 hospitals in the United Kingdom. Most of the nurses filled out the questionnaire during a team meeting in which their manager was present.

Could there be methodological bias that may have affected the research?

1) Yes, social desirability bias.
2) Yes, halo effect.
3) Yes, selection bias.
d) No, there is no clear indication of bias.

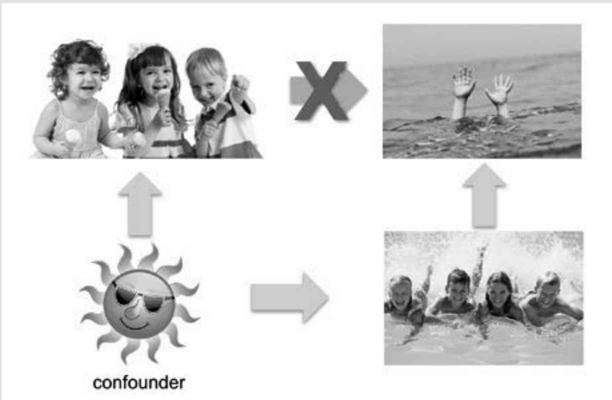
How could the researchers have minimized bias?

a) By picking a truly random sample.
b) By letting the team managers sign a statement of confidentiality
c) By assuring the nurses' anonymity
d) By promising the nurses that their answers will be kept strictly confidential

Limitation 3: Confounders

The third limitation of human judgment that science aims to overcome is confounding: the idea that a third variable distorts (or confounds) a relationship between two other variables. For instance, when factor X seems to causes outcome Y, that relationship could be confounded by a third factor (C) that has a causal influence on both factor X and outcome Y. In that case, C would be an alternative explanation for the observed relationship between X and Y. Sounds confusing? Let's have a look at the following example.

Example



It was long believed that when more ice cream is sold, more children who go swimming will drown, because when children eat ice cream, their stomach gets cold, thus their body reacts by withdrawing blood from their limbs to their abdomen. So, when children go swimming after eating ice cream, their arms and legs are less saturated with blood, causing a shortage of oxygen in their muscles, which causes muscle fatigue, and, as a result, they are more likely to drown. This explanation is, of course, nonsense.

When we take a closer look at the relationship between children eating ice cream (A) and children drowning (B), we will see that this relationship is distorted by a confounder: hot sunny weather (X). When the weather is nice and the sun is shining, more children will eat ice cream, but also more children will go swimming; when more children go swimming, a larger number of children will drown. So, despite the initial explanation of this relationship, eating ice cream (A) does not lead to more children drowning (B); nice, sunny weather (X), however, does.

By applying rigorous research methods, we can prevent confounders from distorting our observations. For various reasons, however, such methods are not always applied, and as a result, research can produce very misleading results. Here are some other examples. Research has shown that

1. at primary schools, children with a large shoe size have better handwriting.
2. people who drink alcohol are more likely to die of lung cancer (the same is true for people who gamble).
3. children who use a nightlight are more likely to develop myopia (nearsightedness).

You probably easily worked out what the confounder is in the first example: age. When children are older, their feet are larger, but because they are older, they have also spent more time practicing writing, so the quality of their handwriting is better. In the second example, the confounder is smoking: people who drink alcohol (or who gamble) are also more likely to smoke, and people who smoke are more likely to die of lung cancer.

The third example is a little more complicated. It was long assumed that children who use a nightlight are more likely to develop nearsightedness. This assumption was investigated in a scientific paper that received a lot of publicity. (13) The findings we e that exposure to nighttime light before the age of two was indeed related to the incidence of myopia. Nearly a year later, other researchers – who used a more rigorous research design – failed to replicate the findings. They did, however, find another result: myopic parents were more likely to leave the lights on at night. The explanation for this phenomenon is simple: myopic parents can't see well at night, and because young children often require nighttime visits, they thus prefer to leave a nightlight on in bedrooms. We know that genetic factors play an important role in the development of myopia; therefore, the relationship between nighttime lighting during early childhood and the later development of myopia is based on a confounding variable: parental myopia. (14)

Learn by doing 6.3

Read the following scenario.

A large international epidemiological study among a random sample of 10,000 students shows that people who wear leather shoes in bed at night often suffer from a headache the morning after.

What is the most plausible explanation for this finding?

a) It is probably just a coincidence.

b) Wearing leather shoes during sleep causes a headache.

c) There is probably a confounder.

d) Due to their unhealthy lifestyle, students suffer more often from headaches in the morning than do other people.

Did I get this 6.6

Read the following scenario.

A large epidemiological study among more than 20,000 factory workers shows that, on average, they have a three times higher risk of lung cancer than the general population. Labor unions and interest groups point out that this is probably caused by the use of workplace chemicals.

Do you agree that chemicals in the workplace is the most plausible cause?

a) Yes, because factory workers are more likely to have jobs that expose them to chemicals, and chemicals are a well-known cause of lung cancer.

b) No, the findings are inconclusive.

Limitation 4: The placebo effect

In medicine, a great deal of research has been done on a phenomenon that for a half-century has been known as the ***placebo effect***: a genuine effect, which is attributable to a patient receiving fake treatment or an inactive substance (e.g., a sugar pill or an injection with distilled water). The treatment has no medical or healing power, so its effect is therefore due to other factors, such as a patient's hope and expectations or trust on a positive outcome. Due to the placebo effect, sugar pills often have the same medical effect as a genuine pill, and even fake operations can sometimes improve a patient's health simply because the person expects that it will be helpful. (15)

The placebo effect is present not only in medical treatments but also in any intervention that aims to influence human behavior, including management interventions. Among notable examples are the Hawthorne experiments, conducted between 1924 and 1933 by Elton Mayo and Fritz Roethlisberger, who examined the relationship between the productivity and working conditions of factory workers. (16) (17)

When the researchers increased the level of light that the workers were subjected to, productivity improved. The same happened when the working conditions were changed in other ways, such as the introduction of rest breaks. However, when the researchers decreased the level of light, the productivity increased even more. In fact, when all working conditions were restored to how they had been before the experiments began, productivity at the factory was at its highest level! What happened was that the placebo effect had affected the outcome: productivity gain did not increase because of the intervention (improving the employees' working condition) but because of a psychological factor: the motivational effect on the workers caused by the researchers showing interest in them. (18) Such a placebo effect can occur in organizational interventions that provide special treatment or attention to the participant and requires careful research design in order to rule it out.

The placebo effect is considered to be a special type of confounder that is present in all scientific studies in which an effect on human beings is involved. As we will see, however, using an appropriate research design that includes a 'fair' comparison can minimize the chance of the placebo effect or other confounders affecting the outcome.

Learn by doing 6.4

If the placebo effect is present in all scientific studies in which an effect on human beings is involved, then it is also present in the assessment of the effect of management interventions that are aimed at the behavior of employees.

Read the following scenario.

A large Midwestern US manufacturing organization decided to deploy a stress-reduction program that includes on-site chair massage therapy, a technique that has successfully been tried by several companies, including AT&T, Apple, and Google.

To determine whether the program is effective, the participants' stress level was measured both before and after its start. The results demonstrated that the program was effective: after several 30-minute massage sessions, participants showed substantially lower stress levels.

The results demonstrated that chair massage therapy results in lower stress levels. But what would be a good method (= 'fair' comparison) to determine whether the reduction of participants' stress level was due to the chair massage rather than the placebo effect? Select the BEST answer.

a) Comparing the stress level of employees who had the chair massage with the stress level of a random sample of employees who did something relaxing for 30 minutes (e.g., listened to music, took a stroll)

b) Comparing the stress levels of the employees measured before the chair massage with the stress levels measured after the chair massage.

c) Comparing the stress level of employees who had the chair massage with the stress level of a random sample of employees who continued their daily work

d) Comparing the stress level of employees who had the chair massage with the stress level of a random sample of employees who took a 30 minute break instead.

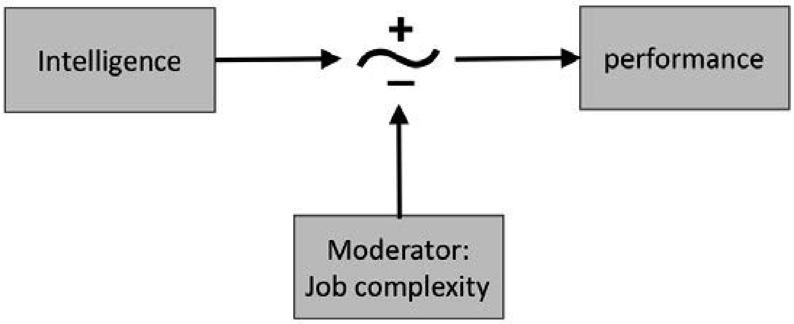
Limitation 5: Moderators and mediators

Finally, another important influence on findings that science seeks to understand is the effect of moderators and mediators. Human beings are very good at detecting the presence of an effect, but they do less well when it comes to identifying factors (either process or contextual) that may have caused or influenced that effect. In fact, in many cases, researchers are less interested in whether something works than in whether there are factors that (positively or negatively) influence the outcome. In science, we refer to these factors as *moderators* and *mediators*.

Moderators

A moderator is a variable that affects the direction and/or strength of the relationship between a predictor (e.g., intelligence) and an outcome (e.g., work performance). Put differently, moderators indicate when or under what conditions a particular effect is likely to be stronger or weaker. A well-known example is the effect of intelligence on job performance. (19)

In general, intelligence is a good predictor of work performance: the higher a person's IQ, the higher his or her performance. However, the extent to which intelligence has an effect on job performance partly depends on the level of job complexity: when a highly intelligent person has to perform relatively simple tasks, his or her performance may not be much different than a less intelligent person's. Conversely, when the task is complex, the highly intelligent person is likely to outperform those of lesser ability. Thus, the positive effect of intelligence on performance is moderated (or depends on) job complexity.



Mediators

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 predictor and the outcome will no longer exist. For example, as early as the 18th century, it was found that eating vegetables prevents sailors from getting scurvy. It was not until the early 20th century, however, that scientists found the specific mediator: vegetables p event scurvy only when they contain vitamin C. A similar example can be found in management. In general, task feedback has a positive effect on a person's performance. This effect, however, is mediated by a person's reaction to the feedback. (20) Put differently, it is a person's reaction to task feedback, and not feedback per se, that determines the extent to which his or her performance will improve.

Learn by doing 6.5

Read the following scenario.

Forty percent of diagnosed breast cancers are detected by women who feel a lump. For this reason, adult women of all ages are encouraged to perform a breast self-exam (BSE) at least once a month. Several medical studies, however, have shown that there is a relation between a woman's socioeconomic status (SES) and the frequency of BSE: the lower a woman's SES, the lower the frequency of BSE. Follow-up studies show a more nuanced picture: the relation between SES and BSE seems to be stronger for older women and less strong or nonexistent for younger women. In addition, it was found that education may have an effect on the relation between SES and BSE. When you remove the effect of education, the relation between SES and BSE disappears!

In this scenario . . .

a) there are no moderators or mediators.

b) age is a mediator and education is a moderator.

c) age is a moderator and education is a mediator.

Did I get this 6.7

Read the following scenario.

Corporate social responsibility (CSR) is a management concept whereby companies integrate social and environmental concerns in their business operations and interactions with their stakeholders. Particularly in the hospitality industry, CSR is a hot topic, in part because a recent study demonstrated a positive relationship between CSR and a hotels' financial performance (HFP): the more a hotel invests in CSR, the higher its HFP. Follow-up studies, however, have suggested that several variables influence this relationship. For example, it was found that a hotel's sales orientation is an important factor: when a hotel focuses only on selling rooms (rather than caring about customers' needs), the relationship between CSR and HFP no longer exists. Thus, when the researchers took this variable out of their model, the relationship disappeared. In addition, it was found that "size of the hotel" also affects the relationship, because small hotels are much more exposed to risks than are large hotels. As a result, larger hotels benefit more from their investment in CSR than do smaller hotels.

In this scenario . . .

a) hotel size is a mediator and sales orientation is a moderator.

b) hotel size is a moderator and sales orientation is a mediator.

c) there are no moderators or mediators.

Qualitative versus quantitative research

Qualitative research is research that uses data that are not expressed in numbers. These data are usually obtained from interviews, focus groups, documentary analysis, narrative analysis, or participant observation. Qualitative research is often exploratory research: we use it to gain an understanding of underlying reasons, opinions, motivations, or mechanisms or to generate hypotheses and/or theories that we can test through quantitative research.

Quantitative research, on the other hand, is research that uses data that are quantified in various ways, that is, measured and expressed using numbers. These data are usually obtained from surveys, tests, financial reports, performance metrics, or statistics. We often use this type of research to generalize results to a larger population, uncover patterns and relationships between variables, or measure the size of an effect of an intervention on an outcome.

Learn by doing 6.6

1. **Read the following abstract of an empirical study, "Understanding the Breast Cancer Experience of Women: A Study of African American, Asian American, Latina and Caucasian Cancer Survivors."**

   Breast cancer is the most common form of cancer in American women across most ethnic groups. Although the psychosocial impact of breast cancer is being studied, there is little information on women from diverse ethnic and socioeconomic backgrounds.

   We conducted a study with breast cancer survivors (BCS) of various ethnicities. A total of 102 BCS participated in focus group interviews (24 African Americans, 34 Asians, 26 Latinas and 18 Caucasians); 20 health professionals participated in key informant interviews. The prevailing concerns among all women included overall health, moderate physical concerns, cancer recurrence or metastases, psychosocial concerns related to worry about children and burdening the family, and body image and sexual health concerns. Additional challenges included: lack of knowledge about breast cancer; medical care issues such as insurance, cost and amount of time spent with physician; cultural sensitivity of providers; language barriers; cultural factors related to beliefs about illness, gender role and family obligations (e.g., self-sacrifice). These BCS, particularly the women of color, voiced that their spiritual beliefs and practices are central to their coping.

This study accomplishes two goals: it adds to the sparse literature concerning the psychosocial sequelae of breast cancer among women of color, and it increases our knowledge of specific cultural influences (e.g., dietary practices, coping) and socio-ecological factors on quality of life (QOL). More importantly, the study addresses areas that have not been studied before, specifically, an in-depth study on BCS QOL comparing multiple ethnic groups in the US. The results of this investigation will provide preliminary information to survivors and health-care providers about the impact of culture and socio-ecological contexts on survivorship.

This is an example of a . . .

a) qualitative study
b) quantitative study
c) empirical study
d) unclear, more information is needed

2. **Read the following abstract of an empirical study, "The Effect of Learning vs. Outcome Goals on Self-Efficacy, Satisfaction and Performance in an MBA Program."**

The present field experiment examined the application of goal setting theory on student self-efficacy, satisfaction with the MBA program, as well as performance (i.e., GPA). Immediately after setting specific high goals, the self-efficacy of MBA students who set year end (distal) outcome goals was lower than participants in either the "do your best" or the learning goal conditions. Participants who set specific difficult learning goals had higher satisfaction with the MBA program than those in other experimental conditions. GPA was significantly higher in the learning goal condition relative to the distal performance goal condition. Participants who set proximal goals, in addition to a distal outcome goal, had a higher GPA than those who only set a distal goal or those who were urged to do their best.

Key measures in this study were satisfaction, self-efficacy, and GPA. GPA was assessed on a 4-point scale, at the end of the academic year. Self-efficacy was measured in terms of obtaining 1/4, 1/2, 3/4, all, or even exceeding one's goals of the MBA program (yes/no) as well as one's confidence in one's ability to do so (10-point scale). At the end of both the first semester and the academic year, a person's satisfaction with the MBA program was assessed using a 12-item scale.

This is an example of a . . .

a) qualitative study
b) quantitative study
c) unclear, more information is needed

Did I get this 6.8

Read the following abstract of an empirical study, "Flexible Work Hours and Other Job Factors in Parental Time with Children."

Flexible working hours are typically seen to be advantageous to working parents, as the flexible hours more easily allow responsibilities of care and employment be balanced. But do flexible work hours actually mean that parents can spend more time with their children? This article explores this for parents of young children in Australia. The analyses use the time use diaries of children in the two cohorts of the Longitudinal Study of Australian Children (LSAC), from the first wave of the study in 2004. The study children in each cohort were aged about one year old and 4–5 years old. For each child, a weekday and weekend diary were completed, giving 5,579 weekday diaries and 4,478 weekend diaries.

The diaries captured details of the children's activities and of who they were with in each 15 min period of a day, and so allowed calculation of the total amount of time the child was with their mother and with their father. Multivariate analyses were used to determine whether amounts of mother-child and father-child time varied according to flexibility of work hours, taking account also of other job characteristics, family and child characteristics. The analyses showed that flexible work hours had only weak independent relationships with mothers' and fathers' time with children. Inasmuch as flexible hours are beneficial for parents, it appears that this is related to their ability to distribute their time between work and family time, rather than giving them more time with children.

This is an example of a . . .

a) qualitative research

b) quantitative research

c) unclear, more information is needed

Quantitative research is widely considered to differ fundamentally from qualitative research. In fact, a common prejudice is that quantitative research is objective and concerns hard data, whereas qualitative research is subjective and concerns soft data. People often take sides and favor qualitative research over quantitative research, or vice versa. A pro-qualitative research person might say: "Quantitative research is useless because it is all about numbers and averages. But people are not numbers, and employees and organizations are never average. If you want to get a better understanding of phenomena in an organization, you should talk to the people who work there, observe closely, and experience what's happening for them. Subjectivity and interaction are key." A pro-quantitative person might respond: "Qualitative research is totally subjective, and thus biased and flawed. You should instead focus only on measurable outcomes, discovering general patterns and explanatory laws that can be translated to the local context. What can't be measured can't be managed, so objectivity and quantitative measures are key."

This is, of course, something of a nonsensical discussion, just as it would be nonsensical to discuss which car is better, a Formula 1 car? A small Fiat? Or a Range Rover? If you wanted to travel fast from A to B and the road was straight, the Formula 1 would likely be the best car. However, if you wanted to drive through the old center of Naples, it would be a bad choice—you would appreciate the small Fiat as you navigated those congested streets. But if you want an off-road adventure in the inlands of Guatemala, you would do well to trade the Fiat for a Range Rover. As you can see, the answer to which car is better always depends on the situation. The same is true when considering which is better: qualitative or quantitative research (or any other type of research): it depends on what you want to know, and hence, it depends on the research question. This principle is also known as *methodological appropriateness*, which we discuss later in more detail.

Research designs

A research design is the “blueprint” of a study that describes its steps, methods, and techniques used to collect, measure, and analyze data. Examples of study designs we frequently use in management are cross-sectional studies, experiments, case studies, and meta-analyses. However, in our field, all kinds of study designs a e used, sometimes with exotic names that make it difficult to fathom exactly which research methods were used. Below we provide a table that describes the elements of common research designs in the social sciences. It is by no means comprehensive, but it provides a basic frame of reference.

Name	Description	Elements
Systematic review	A study that aims to identify as thoroughly as possible all the scientific studies relevant to a particular subject and to assess the validity and quality of the evidence in each study. As the name indicates, a systematic review takes a systematic approach to identifying studies and having their methodological quality critically appraised by multiple researchers. Sometimes includes a *meta-analysis*	Depends on the studies included. Uses a systematic and transparent search process that specifies the criteria used to include or exclude studies.
Meta-analysis	A study that statistically summarizes a large number of studies on the same topic by combining the results of individual studies to get a more accurate estimate of the statistical effect. Sometimes includes a *systematic review*	Depends on the studies included. Averages effect sizes of quantitative studies on a given topic.
Randomized controlled study	A study wherein participants are randomly assigned to a group in which an intervention is carried out (experimental group) and a group in which no (or an alternative) intervention is conducted (control group), and the effect is measured after (and also often before) the intervention. Also referred to as *randomized controlled trial (RCT), experiment, true-experiment*	Random assignment, control group, before-after measurement.
Controlled before-after study	A study wherein participants are (not randomly) assigned to a group in which an intervention is carried out (experimental group) and a group in which no (or an alternative) intervention is conducted (control group), and the effect is measured both before and after the intervention. Also referred to as *CBA, nonrandomized controlled trial (NRCT), quasi-experiment, observational study, controlled longitudinal study, comparison group before-after study, nonequivalent control group design*	Control group, before-after measurement.

Name	Description	Elements
Longitudinal study	A study that involves repeated observations (measurements) of the same variable(s) over a certain period of time. Also referred to as *observational study*	Before-after measurement
Cohort study	A study wherein large groups of participants (also called a cohort or panel) are followed over a long period (prospectively) to see whether differences arise among the groups. Also referred to as *panel study, observational study, longitudinal study*	Control group, before-after measurement (prospective).
Case-control study	A study wherein one group of participants with a particular outcome is compared (retrospectively) with a group that did not experience this outcome. The starting point of the study is the outcome (dependent variable) rather than the intervention or exposure (independent variable). Also referred to as *observational study*	Control group, before-after measurement (retrospective).
Controlled study	A study wherein subjects are (not randomly) assigned to a group in which an intervention is carried out (experimental group) and a group in which no (or an alternative) intervention is conducted (control group), and the effect is measured only after the intervention. Also referred to as *controlled posttest only/comparison group design*	Control group, after measurement.
Before-after study	A study wherein data are obtained or particular variables are measured before and after an intervention, exposure, or event. Also referred to as *single group before-after study* or *longitudinal study*	Before-after measurement.
Posttest only	A study wherein data are obtained or particular variables are measured only after an intervention, exposure, or event. Also referred to as *pre-experimental design, one-shot survey,* or *one-shot case study*	After measurement (prospective).
Cross-sectional study	A study wherein a large number of data or variables is gathered at one point in time, and the intervention or exposure (independent variable) and outcome (dependent variable) are measured simultaneously. It provides a snapshot of the current situation. Also referred to as *survey* or *correlational study*	Cross-sectional (retrospective).
Case study	A study wherein a large number of aspects of a single case (organization or team) are investigated in depth over a long period within the case's own context. A case study is often used to narrow down a broad field of research into an easily researchable practical example. It is a useful design when not much is known about an issue or phenomenon. Researchers using a case study design often apply a variety of (often qualitative) methodologies and rely on a variety of information and data sources. Sometimes referred to as *field study*	Often qualitative methods are used.

Name	Description	Elements
Action research	A study carried out during an activity or intervention to improve the methods and approach of the people involved. The research is conducted by (and for) those taking the action: it is typically designed and conducted by practitioners who analyze their own data to improve their practice. Action research follows a characteristic cycle whereby an exploratory stance is adopted to learn by doing. Often, a variety of (qualitative and quantitative) methodologies are applied. Also referred to as *community-based research, participatory action research*, or *collaborative inquiry*	Often a variety of methods are used.
Ethnographic study	A study wherein researchers completely immerse themselves in the lives, culture, context, or situation (e.g., merger between two organizations) that they are studying. The study is designed to explore cultural phenomena where the researcher observes the organization from the point of view of the participants (e.g., employee or manager) in the study. Also referred to as *field research* or *naturalistic inquiry*	Qualitative methods.
Mixed methods study	A study that involves collecting, analyzing, and integrating quantitative (e.g., experiments, surveys) and qualitative (e.g., focus groups, interviews) research. Also referred to as a *triangulation design*	Both qualitative and quantitative methods are used.

Learn by doing 6.7

The table above shows that in the social sciences, many different research designs are used. It is not useful (or necessary) to learn all these designs by heart, especially because several other terms and classifications a e used. Therefore, for this assessment, use the table as a reference.

Read the following four scenarios and choose the corresponding research design.

Scenario 1

A large Midwestern US manufacturing organization decided to deploy a stress-reduction program that includes on-site chair massage therapy, a technique that has successfully been tried by several companies, including AT&T, Apple, and Google. To determine whether the program is effective, the participants' stress level was measured both before and after its start. The results demonstrated that the program was effective: after several 30-minute massage sessions, participants showed substantially lower stress levels.

What is the research design of this study?

a) A before-after study

b) A controlled before-after study

c) A meta-analysis

d) A cross-sectional study

Scenario 2

The researchers of the effects of chair massage therapy decide to repeat the study, but in order to improve its validity they add a control group: participants will be randomly assigned to a group that receives chair massage therapy and a group that will do something relaxing for 30 minutes (e.g., listen to music, take a stroll) instead.

What is the research design of this study?

a) A mixed methods study
b) A controlled before-after study
c) A quantitative study
d) A randomized controlled study

Scenario 3

Five years after the publication of their study on the effects of chair massage therapy, the researchers notice that several studies on the same topic have been published. They therefore decide to identify as thoroughly as possible all (published and unpublished) studies on this topic and to assess the validity and quality of each study. If similar outcome measures are used, they will also try to statistically combine the results of the studies to calculate an overall effect.

What is the research design of this study?

a) A systematic review
b) A cross-sectional study
c) A mixed methods study
d) A controlled before-after study

Scenario 4

Now the researchers have convincingly demonstrated that chair massage therapy has a positive effect on occupational stress. The question remains, though, how employees feel about this intervention and why it has a positive effect. For this reason, the researchers decide to thoroughly examine – for a period of 6 months – a corporation that uses chair massage therapy. During this period, a large number of aspects will be investigated by applying a variety of quantitative (surveys, data analyses) and qualitative (interviews, focus groups) methods.

What is the research design of this study?

a) An ethnographic study
b) Action research
c) A case study
d) A qualitative study

Did I get this 6.9

Scenario 1: **Read the following abstract of an empirical study, "The Influence of a Manager's Own Performance Appraisal on the Evaluation of Others."**

This study examined the possibility that the performance appraisal process is affected by a pervasive and inherent effect that has heretofore been unidentified. This effect derives from the results of the performance appraisal most recently performed on the manager who subsequently conducts appraisals of others. In the current study, the ratings received by two area coordinators in a university department affected their subsequent ratings of faculty. In a simulation, 30 managers received hypothetical feedback regarding their own job performance. The managers subsequently evaluated an employee on videotape.

Managers who received positive feedback about their performance subsequently rated the employee significantly higher than managers who received negative feedback regarding their own performance. This occurred despite the fact that the managers knew the evaluation of them was bogus.

What is the research design of this study?

a) A controlled before-after study
b) A before-after study
c) A controlled study
d) A posttest only

Scenario 2: **Read the following excerpt from a follow-up study that was conducted.**

In this study, the results of actual performance appraisals conducted by managers who had received their own appraisal a year earlier were examined. The participants were 28 managers employed by a manufacturing company for at least 5 years. The participants were randomly assigned to two groups: one receiving positive feedback and one receiving negative feedback. We received the performance appraisal data of the managers' subordinates one month before (T1) and 12 months after (T2) the managers' own appraisal had been conducted. As in the previous study, no one in the organization was aware of the hypothesis that was being tested. The results of this follow-up study are consistent with the view that one's own performance appraisal is related to the subsequent appraisal of one's subordinates.

What is the research design of this study?

a) A before-after study
b) A controlled before-after study
c) A posttest only
d) A randomized controlled study

Scenario 3: **Read the following abstract of an empirical study, "Generational Differences in Personality and Motivation: Do They Exist and What Are the Implications for the Workplace?"**

Purpose – The purpose of this research is to examine whether personality and motivational driver differences exist across three generations of working Australians: Baby Boomers, Gen Xs, and Gen Ys.

Design/methodology/approach – To test for differences across the three generations, data were collected from a sample of 3,535 managers and [HR professionals] who completed the Occupational Personality Questionnaire (OPQ32) and 294 professionals who completed the Motivational Questionnaire (MQ). All participants were employees of moderate to large Australian organisations.

Findings – The results are not supportive of the generational stereotypes that have been pervasive in the management literature and the media. Specifically, few meaningful differences were found between the three generations. Moreover, even when differences have been observed, these have related more to age than generation.

What is the research design of this study?

a) A meta-analysis
b) A controlled study
c) A cross-sectional study
d) A longitudinal study

MANY STUDENTS WONDER...

What about other classifications, such as experimental research and observational studies?

You are correct – in addition to the research designs discussed in this module, several other classifications and dichotomies are used. We provide an overview of the most common:

- **Descriptive versus exploratory versus causal research**
  Descriptive research describes events or states and aims to find out "what is," so we frequently use observational and survey methods to collect the data. *Exploratory* research aims to generate new questions and hypotheses, investigating underlying reasons or gaining a better understanding of a certain topic. Qualitative methods are often used. Causal research aims to discover causal relationships; hence, we often use research designs with a control group and a pre-measure and designs that assess or measure at multiple time points.
- **Prospective versus retrospective research**
  Most studies are *prospective*: first a research question or hypothesis is developed, then a representative sample of participants is selected and a baseline measure obtained, and finally, after an intervention (or exposure to a variable), the data about the participants are analyzed to examine the effect. In *retrospective studies*, researchers investigate an intervention or exposure by looking back at events that have already happened. Such studies can yield important scientific findings without taking a long time following the participants to find out the outcome
- **Experimental versus observational research**
  Experimental research refers to studies in which the researcher manipulates one or more variables and controls the other variables to determine whether there is a causal relationship between the manipulated variable and the outcome. *Observational* research refers to studies in which the researcher merely observes but does not intervene, and the intention is to find associations among the observed data
- **Experimental versus correlational research**
  In this case, *experimental* research refers to studies that are regarded as "true" experiments (i.e., randomized controlled studies) that allow drawing causal conclusions, whereas *correlational* research concerns studies (e.g., non-randomized or non-controlled studies) that allow conclusions only about correlations or associations.
- **Cross-sectional versus longitudinal research**
  Longitudinal research concerns studies that involve repeated observations (measurements) of the same variable(s) over a certain period of time (sometimes even years); examples are cohort studies and interrupted times series. *Cross-sectional* research refers to studies – such as surveys – in which a large number of data or variables is gathered only at one point in time. It provides a snapshot of the current situation.

How to read a research article

Research findings a e often reported on, in summary form, in newspapers and magazines and on TV. However, those sources often provide information that is too limited to critically appraise or judge a study's trustworthiness. To assess the trustworthiness, we need to find and read the original research article.

Non-academics often consider research articles difficult to read. Practitioners and students sometimes complain that research articles are too lengthy and use too much jargon. They also complain that too many of these articles are dedicated to theory. As a result, non-academics tend to lose track of what the authors are saying or lose interest after a few pages. We should realize, however, that the target audience for academic journals is made up of academics, not practitioners or students. The authors therefore assume that their readers are familiar with academic conventions as well as technical jargon. In other words, research articles are pieces of technical writing whose purpose is to communicate research ideas and findings among researchers.

The most common and important mistake to avoid is expecting that reading a research article is the same as reading an interesting short story or newspaper article and that you can therefore read it in the same way. Research articles are written in a particular style that requires a specific approach in order to fully appreciate and understand what you read. They are not typically intended to be read from the beginning to the end in a linear way. They often require jumping back and forth between the different sections, tables, and appendices. For this reason, reading a research article requires specific skills. Since developing skills requires practice, the more you read research articles, the better you will get at it. To quickly get the most important elements out of a research article, try the following approach:

1. **Step 1. Read the abstract.**
   Start by trying to find out what the article is generally about. What claim(s) is it making? Sometimes the abstract is unclear, but often it provides a summary of the research question, the design, key methods, results and conclusions of the study.
2. **Step 2. Skim the introduction.**
   Quickly skim the first few sections of the introduction, then skim the theory and discussion of previous research.
3. **Step 3. Skim the middle section.**
   Skim the subtitles. Find out what the research questions or hypotheses are. What is the purpose of the research? What are the researchers trying to find out and why
4. **Step 4. Read the method section.**
   Find out what kind of research design the authors used. Did they use a control group and was there a before and after measurement? Try to find which variables and/or outcomes they measured and see if valid and reliable measurement methods were used.
5. **Step 5. Skim the outcome or results section.**
   Skip all elaborate statistical information and focus only on outcomes that are expressed in frequencies, percentages, correlations or other effect sizes. Sometimes the main outcomes are summarized in the first part of the Results. Mo e often outcomes are reported separately for each research question or hypothesis.

6. **Step 6. Read the conclusion.**

   Find out what the authors consider to be the most important outcomes of their study. Are their conclusions justified by the data they presented? Remember that the conclusions authors reach about their research and what it means may not be the same as the conclusions others would draw from the their study.

Skimming = Reading something quickly so as to note only the important points.

In general, don't let yourself be taken in by scientific jargon or complex use of language! Focus on the research question, study design, possible weaknesses, and most relevant outcomes.

Did I get this 6.10

Read the following scenario.

Joanna is interested in the benefits of flexible working. After a search in several research databases, she finds a 32-page research article called "Flexible Working and Performance: A Systematic Review of the Evidence." When she starts reading, she notices the article's introduction and middle section are very lengthy and are dedicated primarily to theory. In addition, the Results section is full of statistical jargon that she finds hard to understand. After about 40 minutes, when she is almost at the end of the article, she has lost track of what the authors are saying and gives up.

What is the most disadvantageous mistake Joanna has made?

a) She read the article from the beginning to the end instead of focusing only on the research question, study design, and most relevant outcomes.

b) She should have consulted multiple research articles on this topic, not just one.

c) Nothing. Sometimes research articles are just hard to read: they are too lengthy and use too much jargon.

d) She first should have gained knowledge on relevant theory, study designs, and common statistical procedures before trying to read research articles.

We started this module by explaining what the basic purpose of science is: to acquire information that will help us to describe, explain, predict, and control phenomena in the world. A core activity of science is gathering objective, external information. However, even the smartest people can easily be fooled into believing something that is not true, so we need to put some safety checks into place when we acquire and appraise external information. Following are three important safety checks used by scientists:

- **The scientific method:** Scientists (1) ask a question about something they have observed, (2) formulate a hypothesis, (3) test the hypothesis through an experiment or by making systematic observations, (4) collect the data, (5) analyze the data, and (5) draw a conclusion.
- **Organized skepticism (peer review):** The scientific community collectively scrutinizes findings   om a position of distrust: the burden of proof is on the scientist with a novel claim.
- **Replication:** A cornerstone of science, replication is repeating studies to see if the same result is obtained.

We also discussed that science is not about truth or proof. Science is about gathering information and testing assumptions (hypotheses) in ways that allow us to estimate how likely it is that something is true. We can never know for sure. After all, it is always possible that new information will cast serious doubt on a well-established model or theory. In addition, even when something seems very close to being proven, it will still be subject to boundary conditions – it can depend on the situation. Finally, as we learn more and more about a given model or theory, we sometimes discover that our original findings we e not quite correct – or at least not specific enough.

If you dig into scientific findings looking for absolute truth and proof, then you will be disappointed. If you look for evidence about likelihoods and probabilities, then research findings can be very useful in overcoming the limitations of human judgment. In this module, we described five of these limitations in detail.

1. **Coincidence:** Could our observation of a phenomenon be due simply to chance? In this section, we discussed scientific concepts such as statistical significance, practical relevance, and p-value. We made the key point that statistically significant research outcomes are sometimes insignificant from a practical perspective (and vice versa).
2. **Methodological bias:** Could our observation be due to personal preference, prejudice, or errors in judgment? In this section, we discussed well-known methodological biases such as selection bias, social desirability bias, and halo effect.
3. **Confounders:** A third variable that distorts (or confounds) a relationship between two other variables is called a confounder. A well-known example is the correlation between the amount of ice cream sold and the number of children who drown during swimming. By applying rigorous research methods, we can prevent confounders from distorting our observations.
4. **The placebo effect:** This effect is present not only in medical treatments but also in any intervention that aims to influence human behavior including management interventions.

5. **Moderators and mediators:** A moderator is a variable that affects the direction and/or strength of the relationship between a predictor and an outcome. A mediator is a variable that specifies how or why particular effect or relationship occurs. Thus, if you remove the effect of the mediator, the relationship between the predictor and the outcome will no longer exist.

We then discussed the difference between qualitative and quantitative research. Qualitative research is research that uses data that are not expressed in numbers. These data are usually obtained from interviews, focus groups, or participant observation. Quantitative research is research that uses data that are quantified in various ways, that is, measured and expressed using numbers. These data are usually obtained from surveys, tests, financial reports, or performance metrics. In this section, we explained that quantitative is not better than qualitative research, nor is qualitative better than quantitative: it depends on what you want to know, hence, it depends on the research question.

At the end of this module, we discussed research designs: the blueprint of a study that describes the steps, methods, and techniques used to collect, measure, and analyze data. Examples of study designs we frequently use in management are cross-sectional studies, controlled studies, case studies, and meta-analyses.

Finally, we discussed how to read a research article. The most common and important mistake to avoid is expecting that reading a research article is the same as reading an interesting short story or a newspaper article and that you can therefore read it in the same way. Research articles are written in a particular style that requires a specific approach in order to fully appreciate and understand them. To quickly get the most important elements out of a research article, you should read the abstract, skim the introduction and middle section, read the Method section, skim the Results section, and read the conclusion. The key point is to not let yourself be taken in by scientific jargon or complex use of language! Focus on the research question, study design, possible weaknesses, and most relevant outcomes.

Podcast

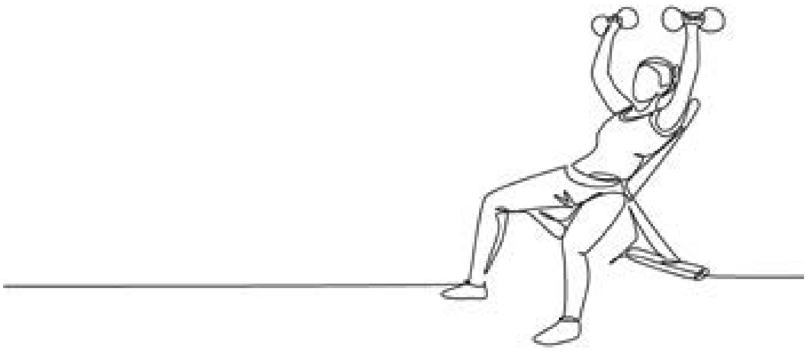
In this podcast host Karen Plum discusses with Eric Barends, Managing Director of CEBMa, Denise Rousseau, Professor at Carnegie Mellon University, and Rob Briner, Professor of Organizational Psychology at Queen Mary University of London the basic principles of scientific research.

In this podcast we look at different aspects of the scientific world – what motivates academics to study the topics they research; the pros and cons of lab and field studies; and how to approach academic studies to get the most from them (don't try to read them cover to cover!). We also discuss the importance of statistical significance and effect sizes in research and their practical relevance in the real world.

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

Exercises



About CATs and Mini-CATs

What is a CAT?

A CAT is an acronym that stands for 'Critically Appraised Topic'. A critically appraised topic provides a quick and succinct assessment of what is known (and not known) in the scientific literature about a claim, assumption, or practical issue by using a systematic methodology to search and evaluate studies. However, in order to be quick, a CAT makes concessions in relation to the breadth, depth and comprehensiveness of the search:

- **Searching:** a limited number of databases may be consulted, and unpublished research is excluded.
- **Research design:** a CAT may be limited to only meta-analyses.
- **Data Extraction:** only a limited amount of key data may be extracted, such as year, population, sector, research design, main findings, and effect size.
- **Critical Appraisal:** quality appraisal may be limited to methodological appropriateness.

Due to these limitations, a CAT is more prone to bias than other types of evidence summaries, such as a systematic review (SR) or a rapid evidence assessment (REA).

What is a mini-CAT?

Given the fact that this is the first time you conduct a CAT, you first need to gain some experience with searching, appraising, and summarizing evidence from the scientific literature. For this reason, you first conduct a 'mini' CAT: a CAT for which we provide the topic, which requires only a limited search, that focusses on meta-analyses, and typically includes only 2 to 3 papers.

Exercise 6.1: Mini-CAT

Read (skim) the Forbes article 'How To Manage Generational Differences In The Workplace' (***download a copy here***) and answer the questions below.

1. The article is based on a scientific study. At the top of the second page it is stated that this study " … involved phone interviews with CFOs, examined workplace differences among different generations, with respect to key business attributes." How would you judge the trustworthiness of this method to examine generational differences?
2. In the next sentence it is stated that *"CFOs were asked: 'In which one of the following areas do you see the greatest differences among your company's employees who are from different generations?"* What do you think of this question?
3. Of course, an important question is "What is known in the scientific literature about (assumed) generational differences in the workplace?". Conduct a mini-CAT to answer this question in which you describe the following items
   a. Your search strategy (search terms, search queries) and search results.
   b. The key features of the included studies (year of publication, population, sample size, study design, main findings). ***Use this matrix***.
   c. A summary of the most important findings.
2. What is – based on the findings of your mini-CAT – your conclusion regarding the claims made in the Forbes article?

Conduct your search in Google Scholar. You can limit your mini-CAT to three studies – one should be a meta-analysis.

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Suggestions for further reading

- *Why we should trust scientists*, Naomi Oreskes, TED Talk, June, 2014
- *The Role of Scientific Findings in Evidence-Based HR*, Briner & Barends, People & Strategy, Spring 2016.
- *Scientific Studies*, John Oliver, Last Week Tonight, May 2016
- *What Science Is — and How and Why It Works*, Neil deGrasse Tyson, Huffington Post, 18 November 201
- *How do placebos work? The science of mind over body*, Jo Marchant, Guardian Science podcast, March 25, 2016

References

1. deGrasse Tyson, N. *What science is—and how and why it works*. Huffington Post. (blog post, November 18, 2016). https://www.huffpost.com/entry/what-science-is-and-how-and-why-it-works_b_8595642
2. Oreskes, N. *Why we should trust scientists.* TED Radio Hour, (June 25, 2014).
3. Brainard, J. (2012). *Physical science concepts for middle school.* CK–12 Foundation. This example is adapted from 1.23, Replication in science. https://www.ck12.org/book/ck-12-physical-science-concepts-for-middle-school/section/1.23/.
4. Wakefield A., Mu ch, S. H., Anthony, A., et al. (1998). "Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children." *The Lancet.* Volume 351. Number 9103. 637-642. [RETRACTED 2010].
5. Deer, B. (2011). "How the case against the MMR vaccine was fixed." *British Medical Journal.* Volume 342. Number 5347. 77-82.
6. For example, when employees must first acquire requisite knowledge or skills to perform the task, specifi and challenging goals can have a negative effect on performance.
7. Allen, N. J., & Meyer, J. P. (1990). "The measurement and antecedents of affective, continuance and normative commitment to the organization." *Journal of Occupational and Organizational Psychology.* Volume 63. Number 1. 1-18.
8. Fisher, R. A. (1973). *Statistical methods for research workers.* Oliver and Boyd. Edinburgh, UK. Edition 14.
9. Neyman, J., & Pearson, E. S. (1933). "The testing of statistical hypotheses in relation to probabilities a priori." *Mathematical Proceedings of the Cambridge Philosophical Society.* Volume 29. Number 4. 492-510.
10. Unfortunately, this interpretation is not completely accurate. You may be surprised, because this is probably what you learned in school or read in textbooks. So, what does a p-value or the term "significant mean? It is very hard to explain what it really means, but this should give you the general idea: the p-value is the probability of your data (e.g., the outcome of a questionnaire), given your hypothesis, the sample population, the sample size, and the statistical method you have used. Still confused? Don't worry about it. As you can see in this video, not even scientists can easily explain what a p-value really means: https://fivethirtyeight.com/featu es/not-even-scientists-can-easily-explain-p-values.
11. Kühberger, A., Fritz, A., Lermer, E., & Scherndl, T. (2015). "The significance fallacy in inferential statistics." *BMC Research Notes.* Volume 8. Number 1. 84.
12. Rosenzweig, P. (2007). "Misunderstanding the nature of company performance: The halo effect and other business delusions." *California Management Review.* Volume 49. Number 4. 6-20.
13. Quinn, G. E., Shin, C. H., Maguire, M. G., & Stone, R. A. (1999). "Myopia and ambient lighting at night." *Nature*. Volume 399. Number 6732. 113.
14. Andrade, C. (2007). "Confounding." *Indian Journal of Psychiatry.* Volume 49. Number 2. 129-131.
15. Stewart-Williams, S., & Podd, J. (2004). "The placebo effect: dissolving the expectancy versus conditioning debate." Psychological bulletin. Volume 130. Number 2.
16. Mayo, E. (1933). The human problems of an industrial civilisation. MacMillan. New York.
17. For a more complete discussion of the Hawthorne effect, see Olson, R., Verley, J., Santos, L., & Salas, C. (2004). What we teach students about the Hawthorne studies: A review of content within a sample of introductory IO and OB textbooks. The Industrial-Organizational Psychologist, 41(3), 23-39.

18. It should be noted that while the Hawthorne effect is plausible more recent research has described the Hawthorne effect found in these original studies as something of a myth and a reanalysis of the data from one of the original studies shows that the effects were not as stong or straightforward as was claimed. See for example: Kompier, M. A. (2006). The" Hawthorne effect" is a myth, but what keeps the story going?. Scandinavian journal of work, environment & health, 402-412. Levitt, S. D., & List, J. A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. American Economic Journal: Applied Economics, 3(1), 224-38.

19. Schmidt, F. L., & Hunter, J. (2004). "General mental ability in the world of work: Occupational attainment and job performance." *Journal of Personality and Social Psychology.* Volume 86. Number 1. 162-173.

20. Chartered Institute of Personnel and Development (CIPD). (2016). *Could do better? Assessing what works in performance management.* https://www.cipd.co.uk/knowledge/fundamentals/people/performance/what-works-in-performance-management-report.

Learn by doing & Did I get this? Answers and Feedback

Did I get this? 6.1

1: A

2: C

3: F

4: B

5: E

6: D

Did I get this? 6.2

a: Correct

b: Incorrect

c: Correct

d: Incorrect

e: Correct

Learn by doing 6.1

a: Incorrect

b: Correct > You probably should. In fact, even if the p-value were 0.20 or 0.30, the odds of making a profit inc ease of half a million dollars might be too good to ignore. What we can learn from this is that the scientific th esholds of 0.05 or even 0.01 may sometimes be too stringent from a practical perspective.

c: Incorrect > More information should not affect your decision.

Did I get this? 6.3

D: Correct > All three answers should raise red flags that learning styles might not be a scientifically supported theory. In fact, meta-analyses have shown that matching instruction to a student’s preferred learning style does not improve learning outcomes. Moreover, a letter titled “No evidence to back idea of learning styles”, signed by 30 academics from the fields of neuroscience, education, and psychology, was published in The Guardian in 2017. You can find the letter here >

Did I get this? 6.4

a: Correct > You probably should not. A p-value of 0.0001 sounds impressive. However, a 0.12 percent lower absenteeism rate is practically irrelevant. Even if a company had a large number of employees, a decrease of 0.12 would not be worth the effort, especially when the cost of implementing the model would be substantial.

b: Incorrect

c: Incorrect > More information would not affect your decision

Learn by doing 6.2

1: Not quite right > You are correct that the sample of 16 companies may have been selected in such a way that some companies are less likely than others to be included, but there is another bias that may have affected the outcome even more.

2: Incorrect > Social desirability bias occurs when research participants answer questions in ways that they think are most socially acceptable. That is not likely to be the case here.

3: Incorrect

4: Correct > The halo effect is a methodological bias in which researchers make generalized inferences based on a few pieces of information. In this case, the reviewers' judgment of whether a company deployed TM practices may be affected by the information about a company's financial performance

a: Not quite right > You are correct that a company's size and annual turnover may induce halo effect, but in this case, there are other characteristics that may have induced halo effect even more.

b: Incorrect > Picking a random sample is a way to minimize selection bias.

c: Incorrect > Assuring the participants' confidentiality or anonymity is a way to minimize social desirability bias.

d: Correct

Learn by doing 6.3

a: Not quite right > You are correct that it may be a coincidence, but given that the findings esult from a large international study that used a large random sample, there is a more plausible explanation.

b: Incorrect > It is not likely that wearing leather shoes at night causes headaches, so there must be a confounder. For example, heavy drinking is associated with forgetting to take off your shoes before going to bed, and as we all know, heavy drinking causes headaches in the morning.

c: Correct > It is not likely that wearing leather shoes at night causes headaches, so there must be a confounder. For example, heavy drinking is associated with forgetting to take off your shoes before going to bed, and as we all know, heavy drinking causes headaches in the morning.

d: Not quite right. You are correct that an unhealthy lifestyle may cause headaches in the morning, but there is a more plausible explanation.

Did I get this? 6.5

1: Correct > Due to the presence of their team manager, the nurses may have been inclined to provide socially desirable answers.

2: Incorrect > The halo effect is a methodological bias in which researchers make generalized inferences based on a few pieces of information. That is not likely to be the case here.

3: Incorrect > Selection bias occurs when participants may have been selected in such a way that some participants are less likely than others to be included. Given that a random sample was used, that is not likely to be the case here.

4: Incorrect

a: Incorrect > Picking a random sample is a way to minimize selection bias.

b: Incorrect > Letting the team managers sign a statement of confidentiality won t minimize the nurses' inclination to provide socially desirable answers.

c: Correct > Letting the nurses fill out the questionnai e anonymously – for example, through an online survey with an anonymous login – would have minimized the risk of social desirability bias.

d: Incorrect > Promising the nurses that their answers will be kept strictly confidential won t minimize the nurses' inclination to provide socially desirable answers.

Did I get this? 6.6

a: Incorrect

b: Correct > A difficulty in investigating occupational haza ds is the problem of disentangling the effects of workplace chemicals from the effects of lifestyle choices, such as smoking habits. You are correct that chemicals in the workplace are a suspected cause of lung cancer. However, smoking also causes lung cancer. So, factory workers are both more likely to smoke and more likely to have jobs that expose them to chemicals. This means that the outcome of this study is inconclusive unless it could rule out smoking as a possible confounder.

Learn by doing 6.4

a: Correct > To determine whether the outcome of the study is affected due to the placebo effect you need a 'fair' comparison, such as a comparison with employees who did something relaxing. Only in that case you would be able to determine whether the lower stress levels were the result of the chair massage therapy, or the result of merely doing something relaxing.

b: Incorrect > A pre- and post measurement is a good way to determine whether there was an effect, but not a good method to determine whether the effect was due to a placebo effect. For that we need a 'fair' comparison.

c: Incorrect > To determine whether the outcome of the study is affected due to the placebo effect you need a 'fair' comparison. Employees who continued their daily work is not a fair comparison, as it is not likely that continue working reduces stress.

d: Not quite right > To determine whether the outcome of the study is affected due to the placebo effect you need a 'fair' comparison. Employees who took a break is not a fair comparison, because we don't know what the employees did during their break (maybe it was something that causes stress).

Learn by doing 6.5

a: Incorrect > Both age and education affect the relationship between SES and BSE, so they moderate or mediate the effect.

b: Incorrect > If you remove the effect of age, the relationship between the SES and BSE still exists, so age is not a mediator. In addition, if you remove the effect of education, the relationship between the SES and BSE disappears, so education is not a moderator.

c: Correct > Age affects the strength of the relationship between SES and BSE, so age is a moderator. In addition, if you remove the effect of education, the relationship between the SES and BSE no longer exists, so education is a mediator.

Did I get this? 6.7

a: Incorrect > If you remove the effect of hotel size, the relationship between the CSR and HFP still exists, so hotel size is not a mediator. In addition, if you remove the effect of sales orientation, the relationship between the CSR and HFP no longer exists, so sales orientation is not a moderator.

b: Correct > Hotel size affects the strength of the relationship between CSR and HFP, so hotel size is a moderator. In addition, if you remove the effect of sales orientation, the relationship between the CSR and HFP no longer exists, so sales orientation is a mediator.

c: Incorrect > Both sales orientation and hotel size affect the relationship between CSR and HFP, so they moderate or mediate the effect.

Learn by doing 6.6

1a: Correct > This research uses data that are not expressed in numbers. The data are obtained from focus groups and interviews. In addition, the study is exploratory in nature: its aim is to gain insight into the breast cancer experience of American women across different ethnic groups.

1b: Incorrect > The study does not use data that are quantified, that is, measu ed and expressed using numbers.

1c: Not quite right > You are correct that this is an empirical study, because it is based on observed and measured phenomena and derives knowledge from actual experience rather than from theory. However, there is a more precise answer.

1d: Incorrect > The abstract clearly states that the study uses data that are obtained from focus groups and interviews. In addition, it states that its aim is to gain insight into the breast cancer experience of American women across different ethnic groups. Based on this information, you should be able to determine what kind of study it is.

2a: Incorrect > This study does not use data that are not expressed in numbers.

2b: Correct > This study uses data that are expressed in numbers. The data are obtained from measurement tools that use a 4-point scale, a 10-point scale, a 12-item scale, and proportions of goal achievement.

2c: Incorrect > The abstract clearly states that the study uses data that are obtained from measurement tools that use a 4-point scale, a 10-point scale, a 12-item scale, and proportions of goal achievement. Based on this information, you should be able to determine what kind of study it is.

Did I get this? 6.8

a: Incorrect > This study does not use data that are not expressed in numbers.

b: Correct > Although the data are obtained from diaries of children, the outcome measure is the total amount of time the child was with their mother and with their father, which is expressed in a number.

c: Incorrect

Learn by doing 6.7

Scenario 1

a: Correct

b: Incorrect > You are correct that the effect is measured both before and after the intervention (chair massage therapy), but the participants were not assigned to an intervention group and a control group.

c: Incorrect > A meta-analysis is a study that summarizes a large number of studies on the same topic. In this scenario, there is only one study.

d: Incorrect > A cross-sectional study is a type of research in which a large number of variables is gathered in one point in time, usually by means of a survey. In this case, the data (level of stress) were collected both before and after the intervention (chair massage therapy).

Scenario 2

a: Incorrect > A mixed methods study is a study in which quantitative methods (e.g., a prepost measure or a survey) is combined with qualitative research (e.g., focus groups or interviews).

b: Incorrect > You are correct that this is a controlled study, but the fact that the participants were randomly assigned to an intervention group and a control group makes it a randomized controlled study.

c: Incorrect > You are right that this is a quantitative study, but there is a more precise answer.

d: Correct

Scenario 3

a: Correct > More precisely, this is a systematic review that includes a meta-analysis.

b: Incorrect > You are right that a large number of data are gathered at one point in time, but in this case, the data concern a large number of studies.

c: Incorrect > A mixed methods study concerns a single study, whereas in this case, the outcome of multiple studies on the same topic are examined.

d: Incorrect > A controlled before-after study concerns a single study, whereas in this case, the outcome of multiple studies on the same topic are examined.

Scenario 4

a: Incorrect > An ethnographic study is a type of (qualitative) research in which the researchers completely immerse themselves in a specific context or situation and observe a phenomenon (in this case, chair massage therapy) from the point of view of the participants (in this case, the employees who receive chair massage therapy). In this scenario, however, a large number of aspects of a single organization is measured, and both qualitative and quantitative methods are used.

b: Incorrect > Action research is a study carried out during an intervention to improve the methods and approach of the participants. In this scenario, the study is not aimed at improving the method (chair massage therapy), but at getting a better understanding of how participants feel about this intervention and why it has a positive effect. In addition, action research is typically designed and conducted by practitioners. In this scenario, the research is conducted by researchers.

c: Correct

d: Not quite right > You are correct that qualitative research methods are used, but there is a more precise answer.

Did I get this? 6.9

Scenario 1

a: Incorrect > Although the ratings of managers who received positive feedback was compared to the ratings of those who received negative feedback, the managers were not assigned (in advance) to two separate groups. For this reason, the study cannot be considered a controlled study. In addition, in a before-after study, the researchers would have collected data on how the managers rate their employees both before and after they received feedback about their own performance. In this case, however, the ratings were collected only *after* the feedback.

b: Incorrect > In a before-after study, the researchers would have collected data on how the managers rate their employees both before and after they received feedback about their own performance. In this case, however, the ratings were collected only *after* the feedback.

c: Incorrect > Although the ratings of managers who received positive feedback was compared to the ratings of those who received negative feedback, the managers were not assigned (in advance) to two separate groups. For this reason, the study cannot be considered a controlled study.

d: Correct > In this case, the researchers collected only data on how the managers rate their employees *after* they received feedback about their own performance.

Scenario 2

a: Incorrect > You are right that the outcome (rating of subordinates) was measured both before and after the intervention (the managers' own performance appraisal), but the managers were also randomly assigned to two groups.

b: Not quite right > You are correct that this is a controlled before-after study, but the fact that the managers were randomly assigned to two groups makes it a randomized controlled study.

c: Incorrect > In a posttest only study, the researchers would have collected the ratings of the subordinates' only after the managers received feedback about their own performance. In this case, however, the ratings were collected both before and after the feedback.

d: Correct > In this case, the managers were randomly assigned to an intervention group (negative feedback) and control/comparison group (positive feedback), and the outcome (rating of subordinates) was measured both before and after the intervention.

Scenario 3

a: Incorrect > This research concerns the investigation of multiple persons, not multiple studies.

b: Incorrect > The data were collected from two samples of two different populations (managers and employees), but these two samples cannot be considered an intervention group and a control group.

c: Correct > The data were collected at only one point/period in time.

d: Incorrect > The data were collected at only one point/period in time.

Did I get this? 6.10

a: Correct

b: Incorrect > Consulting multiple research articles on a topic is often a good thing, but that is not the most disadvantageous mistake Joanna has made.

c: Incorrect > You are correct that research articles are sometimes hard to read, but there is something that Joanna could have done to avoid losing track of the article's meaning.

d: Incorrect > Obviously, having knowledge of theory, designs, and statistics helps, but that is not the most disadvantageous mistake Joanna has made.

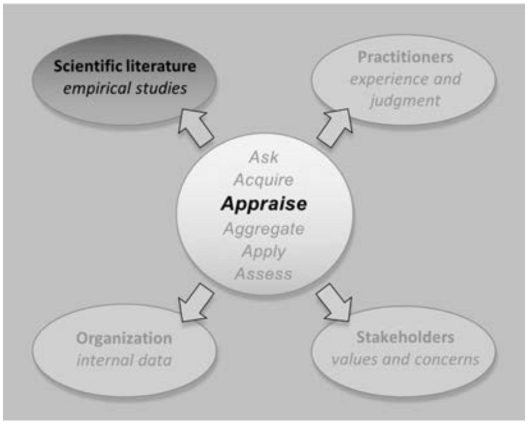
MODULE 7 | APPRAISE Evidence from the Scientific Literature

Learning objectives:

- Determine whether a study's findings are practically relevant
- Determine the precision of a finding or outcome.
- Determine whether an outcome was measured in a reliable way.
- Determine a study's methodological appropriateness.
- Determine a study's methodological quality.
- Grade the trustworthiness of a study.

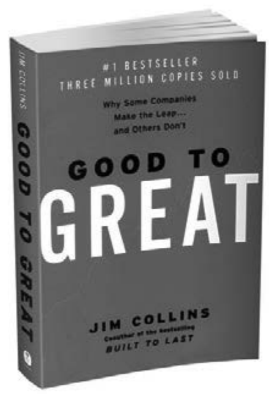
Introduction

> *To be scientifically literate is to empower yourself to know when someone else is full of shit.*
>
> *– NEIL DEGRASSE TYSON*



In 2001, Jim Collins, a former faculty member at Stanford University's Business School, published *Good to Great*. (1) In this book, Collins claims to reveal the principles behind company growth, financial performance, and shareholder value. The book focuses on eleven companies that were just "good" and that transformed themselves into "great" performing companies – where ***great*** is defined as a sustained period of 15 years over which the cumulative total stock return dramatically outperformed the general market and the company's competitors. It took Collins five years to establish how good companies became great. In this period, he and his team of researchers reviewed annual reports, company records, and financial analyses for each company, totaling 980 combined years of data; they conducted 84 interviews with senior managers and board members; and they scrutinized the personal and professional records of 56 CEOs.

Collins's research indicated that every great company had at the helm an ambitious leader combining personal humility with professional will – referred to by Collins as "level 5 leadership." In addition, his research found seven principles that make great companies different from good or average ones. All principles are described in detail in his book, in which he presents them as a framework for success.



The book became a bestseller, selling 4 million copies and going far beyond the traditional audience of business books. Because the principles of *Good to Great* were based on research, they were considered highly trustworthy, and a large number of managers, business leaders, executives, policymakers, hospital administrators, and CEOs used the framework as a blueprint for their own organization.

But over time, what happened with the companies praised in *Good to Great?* You might not be surprised to learn that quite a few have since fallen from grace. In fact, within a few years after the book's publication, all the "great" companies fell into decline. Some of them even filed for bankruptcy, and one company, Fannie Mae, was bailed out by the US government. Pointedly, if you had bought Fannie Mae stock around the time that *Good to Great* was published, you would have lost over 80 percent of your initial investment.

So, what can we learn from this? The obvious answer might be that there is little value in business books that claim to contain a recipe for dramatic success. The less-obvious answer is that, when meticulously collecting and analyzing external evidence, even well-trained and experienced people such as a Stanford professor and his team can come to flawed conclusions if they don't follow the methodological rules of science. This module helps you learn how to critically appraise the strength (trustworthiness) of external evidence such as journal articles, business books, newspaper articles, and textbooks based on scientific research.

Before we discuss the two elements that determine a study's trustworthiness – methodological appropriateness and methodological quality – we need to ask three questions. Only when we can answer these questions positively does it make sense to go through the process of critical appraisal.

Question 1: Are the findings of the study of practical relevance?

Academics are scientific practitioners, not management practitioners. In fact, it is often suggested that academics and management practitioners live in very separate worlds and are interested in different aspects of management and organizations. As a result, the research produced by academics is often of limited relevance to the daily practice of managers. This is especially the case for quantitative studies.

As we explained in Module 6, A Short Introduction to Science, a study's outcome being ***statistically significant*** has limited meaning because it doesn't tell us whether that outcome is of ***practical relevance***.

To determine practical relevance, one of the things we need to consider is the effect size. An effect size is a standardized measure of the magnitude (impact) of the effect. The fact that the measure is standardized means that we can compare effects or differences across different studies, regardless of the variables, measurements, or statistical tests used (2). For example, we could compare an effect size based on customer satisfaction in study A to an effect size based on number of customer complaints in study B.

There are many different measures of effect sizes, but most effect sizes can be grouped into one of two families: the difference between groups (also known as the *d family*) and measures of association (also known as the *r family*). (3) To determine the magnitude of an effect found in a study, you can apply Cohen's rules of thumb. (4)

According to Cohen, a ***small*** effect is one that is visible only 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. The following table provides an overview of how this rule of thumb applies to different measures of effect sizes.

Effect size	Small	Medium	Large
Standardized mean difference: d, Δ, g	≤ .20	.50	≥ .80
ANOVA: η^2, ω^2	≤ .01	.06	≥ .14
Chi-square: ω^2	≤ .10	.30	≥ .50
Correlation: r, ρ	≤ .10	.30	≥ .50
Correlation: r^2	≤ .01	.09	≥ .25
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. (5) For example, a *d* 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, we need to relate the effect size directly to the outcome measured and its relevance, importance, and meaning in each specific context

Effect sizes are typically provided in the Results section of a research paper and/or a separate table. Don't let yourself be taken in by the huge amount of numbers and symbols – just scan through the text and see if you can identify one of the effect sizes listed in the table above.

In addition, if you have two studies that use different effect sizes, you can use the table to make a comparison. For example, if the first study finds a difference of d = .20 between the job satisfaction of two groups, and the second study has a difference of η^2 = .01, you can conclude that the differences found in both studies are small. The same is true for effect sizes within a study. Take, for example, the following table below.

Specific Goals: Effect Sizes and Moderators

Variable	k	d	SE	95% CI Low	95% CI High
Overall effect	49	0.56	0.095	0.37	0.75
Goal difficulty					
Easy (1)	8	0.23	0.111	0.01	0.45
Moderate (2)	15	0.53	0.145	0.25	0.81
Difficult (3)	23	0.80	0.180	0.45	1.15
Task complexity					
Low (1–2.75)	22	0.50	0.092	0.32	0.68
Moderate (3–4.75)	21	0.52	0.163	0.20	0.84
High (5–7)	6	0.94	0.440	0.08	1.80

Note. Findings (k = 49) are based on 739 groups, consisting of 2,954 individuals. k = number of effect sizes; CI = confidence interval.

You can see that the overall effect of goal setting on performance is d = .56, which we consider a medium effect. When you look under Goal difficulty, however, you can see that easy goals have a small effect (d = .23), whereas difficult goals have a large effect (d = .80). Both effects are statistically significant, but only the impact of difficult goals is likely to be practically relevant. Therefore, we can conclude that difficult goals have a large effect on performance.

Learn by doing 7.1

Read the following scenario.

A study that examines the moderating effect of feedback on goal setting (compared to the effect of goal setting alone) finds that there is a statistically significant difference. In addition, it was found that task complexity (complex versus simple goals) and source of the feedback (personal versus impersonal) moderate the effect. In the Results section, the following table is provided.

Based on this table, how would you judge the size of the effects? Check the TWO statements that apply.

Overall Effect of Goals Plus Feedback over Goals Alone

Analyses	*k*	*N*	*Observed Mean d*	*Observed Variance*
Overall	16	744	.59	.19
Complex	6	222	.95	.02
Simple	10	522	.43	.18
Personal	9	404	.61	.22
Impersonal	7	340	.56	.15

a) The overall difference between feedback along with goal setting and goal setting alone is medium.

b) The effect of feedback along with setting complex goals is smaller than the effect of feedback along with setting simple goals.

c) The difference between impersonal feedback along with goal setting and goal setting alone is medium

d) The difference between personal feedback along with goal setting and goal setting alone is small.

Learn by doing 7.2

Read the following scenario.

After a few years, the study on the added value of feedback on goal setting is replicated. Because different statistical methods are used, this second study reports η^2 *(pronounced as eta squared), whereas the first study reported* d*'s. The second study found an overall difference between feedback along with goal setting and goal setting alone of* $\eta^2 = .065$*, whereas the first study reported a difference of* $d = .59$*.*

What would your conclusion be, given these different effect sizes? Use the rules of thumb as a guideline.

a) Both studies report a similar outcome: the difference is medium.

b) The studies report contradicting outcomes: the difference in the first study is large and the difference in the replication study is small.

c) The studies report different outcomes: the difference in the first study is medium and the difference in the replication study is small.

d) The two effect sizes are not comparable.

Did I get this 7.1

Read the following scenario.

A study examines whether there is a difference between the sales performance of sales agents who set their own sales goals (self-set goals) and those whose sales goals are set by their manager (assigned goals). To find out, longitudinal data were collected from 143 industrial salespeople and combined with objective data from company records. In the Results section, a table is provided that reports the outcome (β*, also known as beta coefficient) of a multiple regression.*

Based on this table, how would you judge the size of the effect? Check the TWO statements that apply. Use the rules of thumb as a guideline.

Regression Results

Predictor	Performance
Assigned Goals	0.30***
Self-Set Goals	0.33***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one tailed).

a) The effect of both self-set goals and assigned goals on sales performance is medium.

b) The effect of assigned goals is medium, whereas the effect of self-set goals is large.

c) The difference between the sales performance of sales agents who set their own goals and those whose goals are set by their manager is practically irrelevant.

d) Allowing sales agents to set their own sales goals is not a good idea.

e) Rather than assigning goals, managers should encourage sales agents to set their own goals.

Most quantitative studies include a so-called correlation matrix, which is an overview of the correlation coefficients between all of the variables measured in a study. An example is provided in the following table. As you can see, in this study, a significant correlation was found between sales training A, sales training B, sales training C, and sales performance (–.07, .62, and .11 respectively). Note that only the effect of sales training B (r = .62) is of practical relevance.

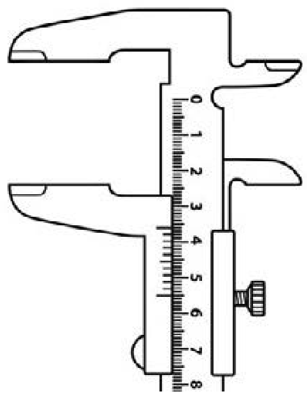
Table 3: Correlation matrix for study variables

No	Variable	1	2	3	4	5	6	7	8	9	10
1	gender	1									
2	age	-.15*	1								
3	education	-.32*	-.02	1							
4	firm size	.03*	.20*	.21	1						
5	firm age	.07	.02	.19	.34	1					
6	sales training A	-.02	.12	.02	.02	.02	1				
7	sales training B	.12	.09*	.04*	.03	.05	.03	1			
8	sales traing C	.09	.48	.01	.22	-.06	.05*	.16	1		
9	experience	.07	.39	.24	.27	-,05	.11*	.05	.32	1	
10	sales performance	.06	.01	.21	.19	-.07	-.07*	.62*	.11*	.13	1

***Significant at 0.05 level**

Obviously, in newspapers, business books, and popular magazines, effect sizes are not normally provided, which makes them an inappropriate source of scientific evidence. However, you may be surprised to learn that many scientific papers published in academic journals also fail to report effect sizes. If this is the case, and no other information is provided regarding the magnitude of the effect or difference found, your conclusion must be that the study provides insufficient information to determine whether it might be relevant for practice.

Question 2: How precise are the findings



Researchers are never able to include all the elements that may be relevant to the outcome of interest in their study. For example, sometimes the population is too large (e.g., surveying all US citizens would take too much time and effort) or that doesn't yet exist (e.g., in the case of a new educational method, we are interested in the population of future students). In those cases, researchers need to rely on a sample – a smaller group that is representative of the whole population.

Effect sizes, however, will differ across samples, meaning that they are only an estimate of the true effect size. The same applies to other parameters, such as mean, percentage, or proportion. So, when an effect size is reported, we also want to know how precise the effect is.

We can determine the precision of the effect size (or any other point estimate) by looking at the confidence interval. A confidence interval provides the upper and lower boundaries between which we expect – usually with a 95 percent certainty – the true value to fall. A confidence interval is stated as 95% CI. If the 95% CI is fairly narrow, then the estimated effect size is a more precise reflection of the true effect size. A 95% CI is considered too wide if the decision you would make based on the value of the lower boundary of the interval would be different from the decision you would make based on the value of the upper one.

Specific Goals: Effect Sizes and Moderators

				95% CI	
Variable	*k*	*d*	*SE*	Low	High
Overall effect	49	0.56	0.095	0.37	0.75
Goal difficulty					
Easy (1)	8	0.23	0.111	0.01	0.45
Moderate (2)	15	0.53	0.145	0.25	0.81
Difficult (3)	23	0.80	0.180	0.45	1.15
Task complexity					
Low (1–2.75)	22	0.50	0.092	0.32	0.68
Moderate (3–4.75)	21	0.52	0.163	0.20	0.84
High (5–7)	6	0.94	0.440	0.08	1.80

Note. Findings (k = 49) are based on 739 groups, consisting of 2,954 individuals. k = number of effect sizes; CI = confidence interval.

When you look at the table above with the effect sizes of goal setting, you can see that the estimated overall effect is *d* = .56, and that the 95% CI is between .37 and .75, meaning that we are 95 percent certain that the true effect size falls somewhere between these two values. This can be considered sufficiently precise. When you look at the effect of easy goals (*d* = .23), however, you can see that the 95% CI is between .01 (which suggests no or only a negligible effect) and .45 (which suggests a medium effect). We consider this confidence interval to be rather wide, because it is less likely that you will decide to implement when the effect is close to zero than when the effect size is near .45.

Learn by doing 7.3

Look at the following table.

Overview of drivers of sales performance

Predictor	Total N	Simple average r[a]	95% Confidence interval Lower bound	Upper bound
Role Perceptions				
Role Conflict	12750	−.11	−.39	.11
Role Ambiguity	27832	−.21*	−.57	−.01
Role Overload	4582	.02	−.37	.42
Burnout	8709	−.15	−.56	.15
Aptitude				
Dispositional Traits	27445	.07	−.31	.48
Personal Concerns	8476	.11	−.16	.45
Identity	26489	.14	−.26	.57
Cognitive	1928	.18*	.01	.45
Skill Level				
Interpersonal	42615	.21	−.12	.65
Degree of Adaptiveness	14547	.26*	.00	.59
Selling-Related Knowledge	29910	.26*	−.01	.67
Motivation				
Cognitive Choice	22989	.15	−.21	.59
Goal Orientation	26460	.18	−.07	.53
Work Engagement	25238	.24	−.11	.67
Personal				
Biographical	44948	.10	−.21	.45
Organizational & Environmental				
External Environment	19506	.17	−.18	.59
Internal Environment	69625	.15	−.17	.54
Supervisory Leadership	49204	.17	−.14	.54

*$p<.05$

This table provides an overview of factors that affect the sales performance of sales agents. For example, when there is role ambiguity (that is, the sales agent feels he or she lacks information to perform his or her job adequately or is uncertain about the company's expectations), the table indicates there is a negative correlation with sales performance.

How would you assess the confidence intervals in this table?

a) The 95% CIs are all too wide.
b) The 95% CIs are sufficiently precise.
c) The 95% CIs are too small.
d) I need more information.

Did I get this 7.2

Read the following finding from a research article, "Virtual Team Leadership: The Effects of Leadership Style on Team Interaction."

To assess Hypothesis 1, which predicted that transactional leadership would more strongly predict a constructive interaction style than would transformational leadership, a 2 × 3 factorial ANOVA was conducted. A significant interaction was found between leadership style on constructive team interaction, $F(2, 57) = 11.22$, $p < .001$, $\eta^2 = .23$ *(95% CI = .18 to .27)*

How would you judge the precision of this finding?

a) The 95% CI is sufficiently precise.

b) The 95% CI is too wide.

c) The 95% CI is too small.

d) The p-value is below .05.

Question 3: How were the findings measured?

The trustworthiness of a study is determined, first and foremost, by the way in which the outcome and other variables are measured. When critically appraising a study, we start by asking, *How was the outcome measured and is that a reliable way of measuring?* In general, the measurement of direct/objective variables (e.g., production errors, staff turnover rate) is more likely to be valid and reliable than that of self-reported/subjective variables (e.g., trust, employee engagement, organizational culture). (6)

In addition, when a study makes use of a questionnaire, you should always check whether the questionnaire has been validated or used in previous studies. In most studies, you can find this information in the Method section, usually under Measurements or Tools.

Did I get this 7.3

In a study on the effect of having employees' dogs present at work, the outcome of interest is occupational stress. Stress can be measured in different ways. Rank order the four measurement methods based on their reliability (the most reliable method on top).

A. **Stress Visual Analog Scale (SVAS)**. Stress on the SVAS is defined as one's response to demanding or unpleasant stimuli or conditions. The scale is 15 cm long and anchored at each end with descriptors of “none” to “the most severe imaginable.” The SVAS is a widely used self-report measure with high levels of compliance and acceptable reliability (Barker et al., 2003).

B. **Impressions from colleagues**. If an employee experiences stress, his or her colleagues will probably notice. This method uses a questionnaire that is filled out by three of the employee's colleagues at the end of the working day.

C. **Stress Observation Scale (SOS)**. Stress on the SOS is defined as the harmful physical an emotional response that occurs when there is a poor match between job demands and the capabilities, resources, or needs of the worker. The scale is filled out by a trained third person who observes whether a participant is stressed.

D. **Salivary Cortisol**. Cortisol is well known to increase with stress and thus is regarded as a highly reliable measure of occupational stress. A single salivary cortisol sample taken at awakening has been shown to be as good as taking multiple samples throughout the day in providing an indicator of overall cortisol production (Yehuda et al., 2003).

Hint: In general, the measurement of direct/objective variables is more likely to be valid and reliable than a measurement based on self-report. In addition, symptoms of stress are hard for a third person to observe.

Did I get this 7.4

Rank order the four outcomes based on how reliable they can be measured (the most reliable outcome on top).

a) The **job attitudes** (e.g., job satisfaction and organizational commitment) of nurses in a local hospital toward performance appraisal

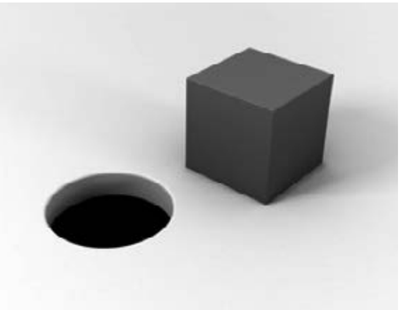
b) The **monthly absenteeism rate** of nurses in a university hospital

c) The **organizational change capacity** of a local hospital anticipating a merger

d) The **number of hip replacements** performed by an orthopedic surgeon per week in a university hospital

If the outcome of a study is not relevant to practice, lacks precision, or is measured in a flawed or highly subjective way, critical appraisal of its trustworthiness is often pointless!

Critical appraisal: Methodological appropriateness



When we find a study that appears to answer our question, we need to know how trustworthy that answer is. How confident can we be that the findings uncovered in the study are valid and reliable? Alternatively, what is the chance that alternative explanations for the observed effect – also known as *confounders*, as we explored in Module 6 – are possible? In other words, to what extent should we trust the findings of a study?

As a rule, a study's trustworthiness is largely determined by its methodological appropriateness. A study's appropriateness is high when its design reflects the best way to answer our question.

Types of Research Questions

In the domain of management, there are types of research questions. These questions can be classified in different ways, but for practitioners, the distinction between cause-and-effect and non-effect questions is perhaps the most relevant:

Cause-and-effect questions

Cause-and-effect questions concern the causal relationship between an action or intervention and an outcome. Examples are:

- Does A have an effect/impact on B?
- Does A work?
- Does A work better than B?

Non-effect questions

Non-effect questions concern aspects other than cause and effect, such as attitude, frequency, or procedure. Examples are:

- How often does A occur?
- How many people prefer A over B?
- Is there a difference between A and B?
- Is A associated with B?
- How do people feel about A?
- Why are people dissatisfied with A

Did I get this 7.5

Read each research question and determine whether it concerns a cause-and-effect question or a non-effect question.

1) What are managers' attitudes toward the relevance and applicability of scientific research findings?
2) What is the impact of a charismatic leadership style on team cohesiveness and overall performance?
3) Nurse absenteeism, stress, and workplace injury: What are the contributing factors, and what can/should be done about it?
4) Human errors in aviation: How often do they occur?
5) Do cultural differences negatively affect the merger between two organizations?

Methodological appropriateness: Cause-and-effect studies (I)

Questions about cause and effect can be difficult to answer. As we will see in this section, controlling the independent variable (the presumed cause) and separating it in time from the dependent variable (the presumed effect) can be very difficult. In addition, many confounding factors may affect the outcome.

This means that if we find a study that appears to answer a cause-and-effect question, we need to make sure that the results of the study are not affected by ***confounders***, alternative factors that might account for the observed effect.

In the realm of science, how well a research design addresses potential confounders is referred to as ***internal validity***. Internal validity indicates the extent to which the results of a study rule out or control for confounding factors or bias and is thus a comment on the degree to which alternative explanations for the outcome found are possible. In the example of the Hawthorne study, which we used in Module 6, *A Short Introduction to Science*, confounding factors affected the outcome. This means that the internal validity of the Hawthorne study and thus the trustworthiness of its findings a e both low. As we will see in the next section, internal validity is a great concern for cause-and-effect questions.

About cause and effect

Many claims, assumptions, and hypotheses in management and leadership are about cause and effect. In fact, cause-and-effect relationships are usually of the greatest interest and relevance to management because a good deal of management-related activity is about making decisions on the basis of expected effects on the organization, its clients or customers, or the people within the organization. Following are some practical examples of cause (A) and effect (B):

- Participative decision-making (A) will lead to higher job satisfaction among our employees (B).
- The Balanced Scorecard (A) is an effective tool to implement our organization's strategy (B).
- Implementing activity-based costing (A) will lead to better financial control of our organization (B).
- If we introduce virtual teams (A), the organization's performance will improve (B).

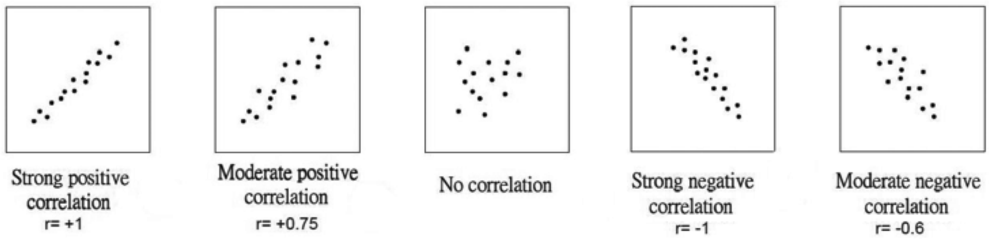
How can we judge the extent to which these claims are likely to be true? Put differently, how sure can we be that it is the cause (A) that leads to the effect or outcome (B) and not some confounder (C) that leads to the outcome? Or will the outcome simply be due to a placebo effect? You probably won't be surprised to learn that this question has preoccupied people since ancient Greece. In fact, it wasn't until the 18th century that the Scottish philosopher David Hume came up with an answer – one that is still used in modern science today. Hume suggested that we should first determine which characteristics a relationship (between A and B) must have before we can call it causal. These characteristics were adapted 200 years later by Paul Lazarsfeld, who described the three criteria for a causal relationship (7):

Three criteria for causation

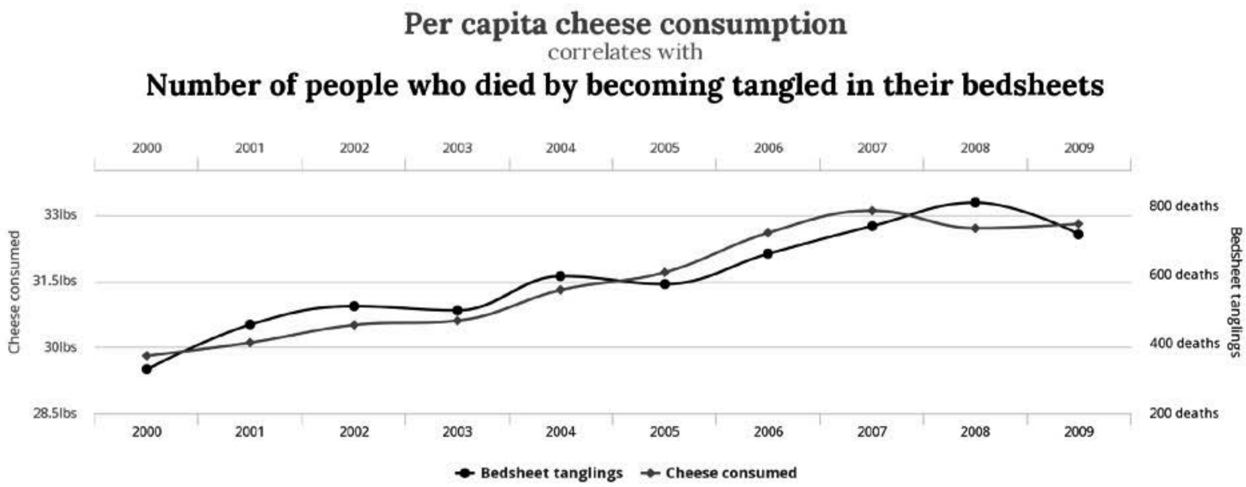
1. **Covariation:** *The two variables A and B are empirically correlated with one another.*
2. **Temporality:** *If variable A is the cause and variable B the effect, then A must occur before B.*
3. **No plausible alternative explanations:** *The observed empirical correlation between A and B cannot be explained as a result of a third variable that causes both A and B.*

Criterion 1. Covariation

Before researchers can demonstrate that a causal relationship exists, they first must show that the e is some relationship at all between two variables. This relationship is referred to as *covariation*: when the value of one variable changes, the value of the other one alters as well. This change may be large or small or even negative (e.g., an increase in one variable leads to a decrease in the other), but there must nevertheless be an empirical (observable, measurable) relationship between the two. Covariation is expressed through a simple metric: a *correlation coefficient* – a numerical index that reflects the strength of the relationship between two variables. (8) The value of this coefficient ranges between –1 and +1. As you can see in the picture below, when the correlation coefficient is high, the points in the fig e are closely aligned, indicating covariation. Conversely, when the correlation coefficient is low, the points in the figure are dispersed, indicating limited covariation. There are different types of correlation coefficients, depending on the type of variables that a e measured, but they can all be considered an *effect size*: a standardized measure of the strength of the relationship. Thus, a correlation of r = .60 is stronger than a correlation of r = .30.



As you have probably heard many times before, however, correlation is not causation. Therefore, when a study describes a correlation between two variables, it only means that the researchers were able to demonstrate "some" relationship. In fact, correlation doesn't even mean that the two variables are directly related. As we saw in the example with the sales of ice cream and the number of children drowning (Module 6, under Limitation 3: Confounders), these two variables closely covary. Moreover, if you were to calculate a correlation coefficient, you would find a strong correlation between the two. It would not matter, either, how many times you measured this relationship: it would always be there. We can learn from this example that the mere fact that two variables correlate does not mean that there is a causal relation, even when the correlation is strong and consistent. In fact, when we measure only whether two variables covary, we can find the most exotic and bizarre relationships. For example, in the graph below, you can see that eating cheese highly correlates with the number of people who died by becoming tangled in their bedsheets! (9)



Nevertheless, covariation is the first criterion for an assumed cause-and-effect relationship. This means that *if* a correlation or effect is found, our next question should be: how was the effect measured (and is that a reliable way to measure it)? This also means that if no correlation or effect is found, we can safely conclude that a cause-and-effect relationship is not likely. (10)

Learn by doing 7.4

Study 1: **Read the following results from a study, "Business-Unit-Level Relationship between Employee Satisfaction and Business Outcomes: A Meta-Analysis":**

Based on 261 cross-sectional studies with a total sample size of 7,939 business units and 21,567 employees, this meta-analysis examined the relationship at the business-unit level between employee satisfaction and performance. We found a significant correlation between overall satisfaction and business-unit performance. Corrected for the performance variable's measurement error and range restriction, the correlation was r = .21

Which two **conclusions can be drawn based on these findings?**

a) The findings suggest that a higher level of employee satisfaction leads to higher performance

b) It is unclear whether a higher level of employee satisfaction leads to higher performance.

c) The findings suggest the e is a weak relationship between employee satisfaction and performance.

d) Given the large number of studies, a causal relation between employee satisfaction and performance is likely.

e) The findings suggest the e is no causal relation between employee satisfaction and performance.

Study 2: **Read the following results from a study entitled "The effects of diversity on business performance: a meta-analysis":**

A meta-analysis of the data from 61 studies was used to examine the impact of two types of diversity attributes, highly job-related and less job-related, on performance. This distinction was used to test the proposition that different types of diversity will differentially impact performance. In addition, type of team was examined as a possible moderator of the relationship between diversity and performance. Results showed that neither type of diversity had a relationship with performance.

What is the best conclusion that can be drawn on the basis of these findings?

a) A cause-and-effect relationship between highly job-related diversity and less job-related diversity and performance is not likely.

b) The findings a e inconclusive.

c) It is unclear whether highly job-related diversity or less job-related diversity leads to higher performance.

d) There is no difference between the impact of highly job-related diversity and less job-related diversity on performance.

Methodological appropriateness: Cause-and-effect studies (II)

Criterion 2: Temporality

When a correlation or effect exists, we then must demonstrate that the cause (A) happened before the effect (B). After all, it could also be the other way around: B causes A. For example, research has demonstrated that married people are, on average, happier than single people. (12)

This suggests that getting married makes you happy. But maybe it is the other way around: if you are a grumpy, unhappy person, you are less likely to get married. For this reason, we need to make sure the effect was not already present before the cause. This means we need a ***baseline measurement***: a calculation of the variables of interest at the beginning of a study that are compared to later measurements to judge the outcome. The baseline measure is usually referred to as the ***before measurement***, whereas the one taken at the outcome is the *after measurement*.

In athletics, for example, a baseline is essential. Without being sure that all athletes in a running event start at the same time, we can't be certain that the winner was indeed the fastest. A baseline is also essential in studies examining a cause-and-effect relationship between two variables. Consider the following example. A study has demonstrated that youngsters who smoke tend to have better lung capacity than youngsters who don't smoke.

You may wonder how that can be. After all, we know that smoking is very damaging to our health, especially our lungs, so how is it possible that young smokers tend to have better lung capacity? Could it be that they compensate for their unhealthy habit with sports and other physical activities? What if we tell you that this study did not have a before measure? This means that we can't rule out that the outcome (better lung capacity) may have been already present before the cause (smoking).

This makes the explanation for this bizarre outcome rather simple: youngsters with a serious lung condition such as asthma are less likely than healthy youngsters to start smoking, so there is an overrepresentation of youngsters with poor lung function among nonsmokers. The lesson here is obvious: when we don't know whether subjects (groups or individuals) are comparable at baseline, we can't say with certainty that there was a cause-and-effect relationship.

Criterion 3: No plausible alternative explanation

Once researchers have demonstrated that the assumed cause (independent variable) and effect (dependent variable) covary and that the effect does not precede the cause in time, they have to rule out plausible alternative explanations. One of the ways to do that is through a study's research design, the blueprint of a study that describes its steps, methods, and techniques used to collect, measure, and analyze data. We already learned that to demonstrate causality, a study must include before and after measurements, but to rule out other explanations for the effect found, researchers must include two additional components in the research design: a control group and random assignment.

Control Group

In a ***controlled*** study, one group (sometimes called the *intervention*, *treatment*, or *experimental* group) that is exposed to a condition or situation expected to have an effect (the assumed cause) is compared with another group (known as the *control* or *comparison* group) that is not. Researchers expose just one group to the variable they expect to have an effect and then compare that group control group to help rule out alternative explanations. In addition, the control group can serve as a benchmark for comparison against the intervention group.

Imagine that researchers want to know whether a new fertilizer makes potted plants grow faster. They make two groups. One group of plants receives the fertilizer (the intervention group) and one group does not (the control group). If they make sure that the lighting, water supply, size of the pot, and all other factors are held constant for every plant in both groups – and thus the only thing that differs between the two groups is the fertilizer given to the plants – only then can the researchers be certain that the new fertilizer causes the difference in growth.

Random Assignment

Even when researchers use a control group, they still can't be 100 percent certain that there were no (unknown) confounders that affected the outcome. For example, in the case above, factors such as lighting, water supply, and size of the pot were equal in both groups. But what if the plants in the control group were obtained from a different grower than the ones in the experimental group? Or what if the plants in the control group were in the left-hand side of the greenhouse and those in the experimental group were on the right? The researchers can't rule out that the plants might be different in terms of nutritional benefits, DNA, or factors of which they are unaware. For this reason, they must use random assignment when creating an intervention and control group.

Random assignment is a method used to create different groups that include subjects (organizations, units, teams, or individuals) with similar characteristics, so that the groups are similar at the start of the study.

The method involves assigning subjects to different groups at random (by chance, like the flip of a coin), so that each subject has an equal chance of being assigned to each group, and any possible distorting factor is equally spread over both groups. Thus, the researchers can more confidently attribute any differences between the groups measured at the end of the study to the variable that is expected to have an effect (the assumed cause)

Random assignment is NOT the same as random sampling!

Random sampling refers to selecting subjects in such a way that they represent a larger population, whereas random assignment concerns assigning subjects to a control group and experimental group in such a way that they are similar at the start of the study. Put differently, random selection ensures high representativeness, whereas random assignment ensures high internal validity. If a study uses the term random or randomized, you should always determine whether it concerns random selection (as in a survey) or random assignment (as in a controlled study).

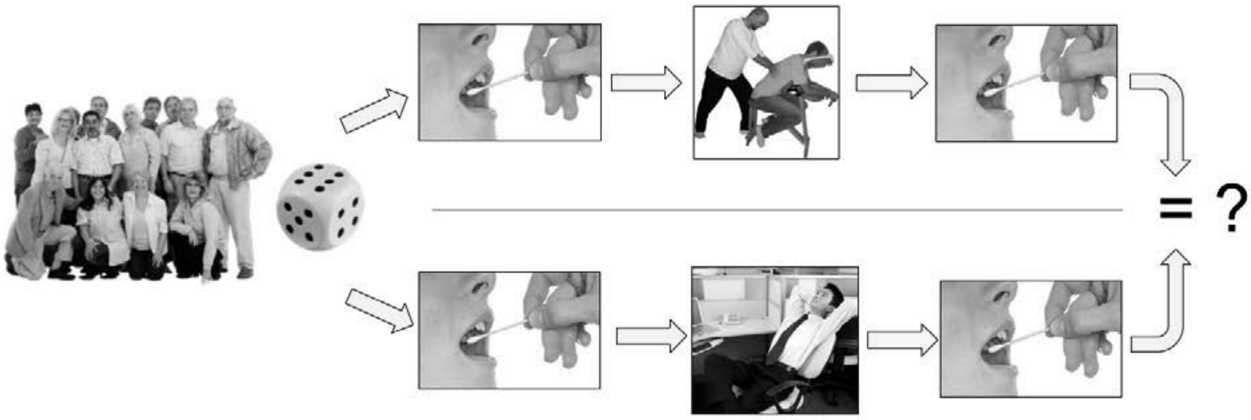
Methodological appropriateness: Cause-and-effect studies (III)

The gold standard of cause-and-effect studies

Based on the elements described above, we can now determine the best – that is, the most appropriate – study to answer a cause-and-effect question with the lowest chance of confounders. Let's consider the example of a researcher who wants to examine the effect of a stress-reduction program using on-site chair massage therapy. The first criterion of causality is to demonstrate, using valid and reliable measurements, that the cause-and-effect correlate. The researcher has already determined that the most valid and reliable way to measure stress reduction would be to measure the cortisol levels (a stress hormone) in the saliva of employees.

The second criterion of causality, temporality, states that the researcher must demonstrate that the cause (chair massage therapy) preceded the effect (stress reduction). The researcher therefore measures the employees' cortisol level both before and after the chair massage therapy. (13)

The third criterion concerns ruling out plausible alternative explanations for the effect found, so the researcher randomly assigns employees to either a control or an intervention group. The researcher first takes a representative sample of the employees and then randomly assigns them to the two groups by flipping a coin. Then the researcher needs to determine what the control group should do to create a valid and reliable benchmark: continue working, take a break, or receive fake massage therapy? After some discussion, the researcher decides that continuing to work is an unfair comparison and fake massage therapy is hard to realize, so the control group will take a break during which they do something to relax. This means that the research design would be as depicted below. This design is known as a *randomized controlled study* or *randomized controlled trial* (RCT), the gold standard to answer cause-and-effect questions.



Don't worry about statistics!

Cause and effect can be established only through the proper research method: valid and reliable measurement methods, before and after measurements, a control group, and random assignment. No amount of statistical handwaving can turn correlations into conclusions about causation!

Learn by doing 7.5

A study's trustworthiness is first determined by its methodological appropriateness. Thus, when it comes to questions about cause and effect, we prefer studies with a research design that is appropriate to answer cause-and-effect questions. As discussed, questions about cause and effect are best answered by a study that meets the three criteria of causation. Based on these criteria, a study's methodological appropriateness can be judged as follows:

Methodological appropriateness	Research design elements
High	1) Measurement of the effect 2) Pre- and post-measurements 3) Control group 4) Random assignment
Moderate	1) Measurement of the effect 2) Pre- and post-measurements 3) Control group
Limited	1) Measurement of the effect 2) Pre- and post-measurements OR 3) Control group
Low	1) Measurement of the effect

Consider this example that was used earlier in the module:

A study examines whether there is a difference between the sales performance of sales agents who set their own sales goals (self-set goals) and those whose sales goals are set by their manager (assigned goals). Sales data were collected from 143 industrial salespeople both before and after the goals were set. Results of a multiple regression indicate that the impact of self-set goals ($\beta = .38$) is larger than the impact of assigned goals ($\beta = .27$).

How would you grade the methodological appropriateness of this study?

a) High
b) Moderate
c) Limited
d) Low

Did I get this 7.6

Read the scenario below.

This study tested hypotheses related to the effect of sales orientation on sales performance. A sample of 621 sales agents of large companies participated in a survey conducted in early 2014. A questionnaire was used for data collection and was sent out to all participants for them to rate their sales orientation: customer orientation (focusing on customer needs) or sales orientation (focusing on the selling of a particular product). Salespersons with high levels of customer orientation showed a higher sales performance than sales agents with a sales orientation ($\beta = 0.06$, $p < 0.01$). This finding suggests that sales orientation has a small but significant effect on sales performance.

How would you grade the methodological appropriateness of this study?

a) High
b) Moderate
c) Limited
d) Low

Did I get this 7.7

Consider this example that was used earlier in the module.

In this study, the results of actual performance appraisals conducted by managers who had received their own appraisal a year earlier were examined. The participants were 28 managers employed by a manufacturing company for at least 5 years. The participants were randomly assigned to two groups: one receiving positive feedback and one receiving negative feedback. We received the performance appraisal data of the managers' subordinates one month before (T1) and 12 months after (T2) the managers' own appraisal had been conducted. As in the previous study, no one in the organization was aware of the hypothesis that was being tested. The results of this follow-up study are consistent with the view that one's own performance appraisal is related to the subsequent appraisal of one's subordinates.

How would you grade the methodological appropriateness of this study?

a) High
b) Moderate
c) Limited
d) Low

Methodological appropriateness: Non-effect studies

Non-effect questions and cause-and-effect questions are answered in different ways. For example, the question *How many hospitals in Denmark have applied Lean Management to their outpatient surgery department?* is best answered by taking a random, representative sample of all Danish hospitals and asking them by phone, email, or letter (i.e., survey research) whether they have applied Lean Management. The same is true of questions regarding nurses' feelings toward the use of Lean Management in their hospitals. In that case, focus-group research could be a good choice. The question of whether Lean Management has a positive effect on the performance of surgical teams, however, can't be answered with a survey or a focus group, because this question is about cause and effect and therefore needs – at minimum – a control group and both before and after measurements.

As discussed in the previous sections, the main concern in cause-and-effect studies is internal validity, the extent to which a study's results may be affected by confounders. In non-effect studies, external validity rather than internal validity is often a larger concern. External validity is the extent to which the results of the study can generally be applied to other situations (e.g., organizations) and to other people (e.g., employees). For instance, in the example above, we want to make sure that our sample of hospitals and nurses accurately represents all Danish hospitals and nurses and that the findings f om our survey and focus group are generalizable to the whole population. This, and other concerns for non-effect studies, means that a randomized controlled trial, the gold standard for cause-and-effect questions, is not an appropriate design for non-effect studies. Following is an overview of common non-effect questions and the corresponding appropriate study designs.

Questions about frequencies

In organizations, questions concerning frequencies are common. In hospitals, for example, common questions would be the number of patients admitted today, how many hip replacements the surgical team performed this week, or how many medical errors were made last month. The same is true of questions regarding the number of employees who are absent, the number of employees who are satisfied with their working conditions, or the number of employees who will retire within the next five years – all common questions for an HR manager. Not surprisingly, many scientific studies also add ess research questions about frequencies. Control groups and premeasures are not needed to answer these questions. The most appropriate research design for answering a frequency question is a cross-sectional survey, as this type of study obtains information on counts.

Questions about differences

Related to frequency questions are questions concerning differences. When the hospital manager has determined how many medical errors were made on average, he or she may also want to know whether there are differences between the medical departments. Likewise, an HR manager may want to know whether satisfaction with working conditions differs among divisions or types of employee. In the scientific literature, differences between organizations, teams, or employees are also often the focus of research. For example, a large number of studies is available on the (assumed) differences between entrepreneurs and managers, knowledge workers and manual laborers, virtual teams and collocated groups, and private companies and public organizations, to name a few.

Usually, these differences concern productivity or performance, but differences regarding attitudinal, motivational, or behavioral factors are also common. Questions concerning differences – as with questions relating to frequency – are best answered with a study that uses a cross-sectional design. Note that a cross-sectional study is a very appropriate design for establishing whether a difference exists, but it does a poor job of determining the cause of this difference. When a difference has occurred over time, a longitudinal (before-and-after) study is the more appropriate design.

Questions about antecedents (predictors)

An antecedent is a variable or factor that precedes a certain outcome. For example, job satisfaction is an important antecedent of employee turnover (14). Thus, when employees' job satisfaction decreases, the likelihood that they eventually will quit their job increases. Research findings about antecedents can therefore be helpful evidence in the decision-making process: they can help identify predictors for a certain outcome relevant to the organization. A longitudinal study – in particular a before-after study – is the most appropriate design to determine whether a variable is an antecedent/predictor for a certain outcome. Note, however, that longitudinal studies are not appropriate to determine whether there is a causal relationship. An antecedent may be a strong predictor for a certain outcome, but it may not necessarily be the cause of that outcome. For example, warm weather is a well known antecedent of the number of crimes in a city: aggressive crime increases when the temperature increases (15). Thus, information from weather forecasts is used to determine where and when the deployment of law enforcement officials and emergency response personnel is most beneficial. However, warm weather does not cause criminal behavior. A more plausible explanation is that when the weather is warm people are more likely to leave their homes and go to public places – this increases social interactions which in turn may lead to more interpersonal conflicts and violence

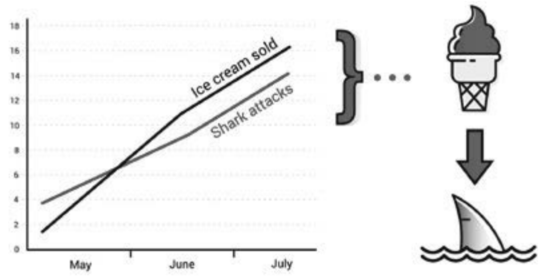
Questions about attitudes and opinions

For some managers, questions about frequencies or differences often relate to attitudes or opinions. In the example of the HR manager, the frequency and possible difference involved an attitudinal factor: employees' satisfaction with their working conditions. Other examples of workplace-related attitudes are organizational commitment, turnover intention, and trust in management. When we have – in advance – a clear idea about the attitude we want to measure, and this attitude is quantifiable, a cross-sectional study is the most appropriate design. If, however, we want to more deeply explore and better understand the attitudes and opinions of people, a study with a qualitative design, such as focus-group research, may be a more appropriate option.

Questions about experiences, perceptions, needs, and feelings

In general, experiences, perceptions, needs, and feelings can be difficult to quantify. For example, the answers to questions such as *How do you feel about the introduction of flexible work arrangements in this company?* or *What do you think you need to optimally function within a virtual team?* can be a challenge to express in numbers. Data relating to experiences, perceptions, needs, and feelings are therefore often obtained from interviews, focus groups, narrative analysis, or participant observation – in other words, qualitative methods.

Questions about associations or relationships



Questions concerning associations fit into a special category. For example, the hospital manager may want to know whether medical errors are related to the time of day or a specific weekday. Determining whether the e is a (co)relation – also called *association* – between two variables is, as discussed earlier, often the first step in identifying a causa relationship. Thus, a cross-sectional study is an appropriate design to determine whether an association exists between two variables. Clearly, study designs that use before and after measurements, a control group, and/or random assignment are also appropriate for measuring the extent of an association, but such designs are used mainly to determine whether the association could be causal.

Learn by doing 7.6

Read the following abstract from a study, "Perceived Neighborhood Safety and Body Mass Index":

We sought to determine whether there is a relationship between perceived neighborhood safety and body mass index (BMI). A random sample of 2,255 adults from the Los Angeles Family and Neighborhood Survey 2000–2001 was analyzed using instrumental variables. The main outcome was BMI using self-reported height and weight, and the main independent variable was residents' report of their neighborhood safety. In adjusted analyses, individuals who perceived their neighborhoods as unsafe had a BMI that was significantly higher than did those who perceived their neighborhoods as safe.

This study aims to answer a question about . . .

a) experiences, perceptions, and feelings

b) frequencies

c) cause and effect

d) association

Learn by doing 7.7

Read the following abstract from a study, "Electronic Health Records: Reports from Early Adopters in Long-Term Care Facilities":

Electronic health records (EHRs) are becoming a required technology across the healthcare sector. Long-term care (LTC) facilities have lagged other settings in adopting health information technologies but represent an area where significant care coordination benefits might be realized. Nevertheless, managers face many of the same challenges implementing EHRs that exist in other environments when implementing enterprise-wide systems. This study was conducted to provide a description of the early users' experiences with EHRs in LTC facilities. Semi-structured interviews were conducted. The 10 sites were all LTC facilities using an EHR as of July 2008 in Texas. The interview respondents included administrators, nursing managers, nurses, certified nurse aides, and other system users. Semi-structured interviews across multiple stakeholders were used to assess constructs critical to EHR adoption and implementation. The employees who work with EHR systems on a daily basis reported mixed experiences.

This study aims to answer a question about . . .

a) cause and effect

b) frequency

c) association

d) experiences, perceptions, and feelings.

Did I get this 7.8

Read the following abstract from a study, "Do Cultural Differences Matter in Mergers and Acquisitions? A Tentative Model and Meta-analytic Examination":

A substantive body of theory and research on the role of culture in mergers and acquisitions (M&A) suggests that cultural differences can create major obstacles to achieving integration benefits. However, the opposite view – that differences in culture between merging firms can be a source of value creation and learning – has also been advanced and empirically supported. In an attempt to reconcile these conflicting perspectives and findings, we present a model that synthesizes our current understanding of the role of culture in M&A, and we develop a set of hypotheses regarding mechanisms through which cultural differences affect M&A performance. The results of a meta-analysis of 46 studies, with a combined sample size of 10,710 M&A, suggest that cultural differences affect sociocultural integration, synergy realization, and shareholder value in different, and sometimes opposing, ways. Moderator analyses reveal that the effects of cultural differences vary depending on the degree of relatedness and the dimensions of cultural differences separating the merging firms, as well as on research design and sample characteristics.

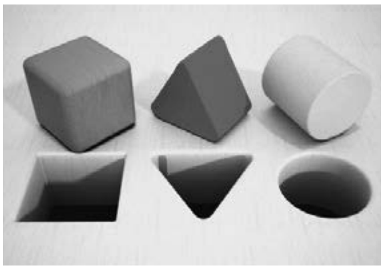
This study aims to answer a question about . . .

a) differences
b) association
c) cause and effect
d) frequencies.

Beware of single studies!

As we explained in Module 6, A Short Introduction to Science, an important safety check to ensure that scientific claims are trustworthy is replication. If novel findings from scientific research can be reproduced, it means they are more likely to be valid. On the other hand, if the findings cannot be replicated, they are likely to be invalid. Therefore, no single study (not even a randomized controlled trial) can be considered to represent strong evidence: a single study can be considered merely indicative. For this reason, we prefer meta-analyses or, even better, systematic reviews. As explained earlier, a systematic review is a summary of studies that aims to identify all relevant studies on a specific topic. A meta-analysis is a summary of studies in which statistical analysis techniques are used to combine the results of individual studies to get a more accurate estimate of the effect. Thus, irrespective of the type of question, meta-analyses and/or systematic reviews are often the most appropriate research designs. In other words, it is the body of evidence – not single studies – that is important.

Methodological appropriateness: Summary



To summarize, when answering questions about cause and effect, we need studies with a research design suited to answering cause-and-effect questions – that is, a study that includes both a control group (preferably randomized) and before and after measurements. When we want to answer non-effect questions, such as questions about associations, frequencies, differences, or attitudes, a cross-sectional survey is better. Finally, when we want to explore feelings, perceptions, needs, or experiences, qualitative methods are likely to be most appropriate.

Following is an overview of the methodological appropriateness of research designs for different types of research questions.

Methodological appropriateness: Which design for which question?

Purpose	Example	RCT	CBA	BA	Contr	Cross	Qual
Effect, impact	Does A have an effect/impact on B? What are the critical success factors for A? What are the factors that affect B?	A	B	C	C	D	na
Prediction	Does A precede B? Does A predict B over time?	A	A	A	D	D	na
Association	Is A related to B? Does A often occur with B? Do A and B covary?	A	A	A	A	A	na
Difference	Is there a difference between A and B?	A	A	A	A	A	na
Frequency	How often does A occur? How many people prefer A?	na	na	na	na	A	na
Attitude, opinion	What is people's attitude toward A? Are people satisfied with A? Do people agree with A?	na	na	na	na	A	C
Experience, perceptions, feelings, needs	What are people's experience with A? What are people's feelings about A? What are people's perceptions about A? Why do people (think they) need to do/use A?	na	na	na	na	B	A
Exploration, theory building	Why does A occur? Why is A different from B? In what context does A occur?	na	na	na	na	B	A

RCT = Randomized controlled trial; CBA = Non-randomized controlled before-after study; BA = Before-after study; Contr = Controlled study; Cross = Cross-sectional study; Qual = Qualitative study; na = not appropriate

When we critically appraise the trustworthiness of a study, we need to consider its methodological appropriateness to the type of question that it aims to answer. Findings from a study with low methodological appropriateness will also have low trustworthiness/certainty, regardless of how well the study was conducted. When critically appraising a study's trustworthiness, we therefore start by identifying its research design. The resultant methodological appropriateness is then expressed in a measure of trustworthiness. Note that we don't judge the trustworthiness of a study as such, because studies are not intrinsically trustworthy or untrustworthy. We can only judge the trustworthiness of a study's findings *given its research design and the type of question it aims to answer.*

Methodological appropriateness	Estimated trustworthiness	Proper wording of the findings
High	90%	It is shown that . . .
Moderate	80%	It is likely that . . .
Limited	70%	It could be that . . .
Low	60%	There are signs that . . .

Meta-analyses and systematic reviews: One level up

As discussed earlier, no single study (not even a randomized controlled trial) can be considered to represent strong evidence: a single study can be considered merely indicative. Therefore, we prefer meta-analyses or, even better, systematic reviews. A systematic review is a summary of studies that aims to identify all relevant studies on a specific topic. A meta-analysis is a summary of studies in which statistical analysis techniques a e used to combine the results of individual studies to get a more accurate estimate of the effect. Thus, meta-analyses and/or systematic reviews have a higher level of trustworthiness than a single study. For this reason, we upgrade a study's trustworthiness by one level if it concerns a systematic review or meta-analysis. For example, when it concerns a cause-and-effect question, a meta-analysis of cross-sectional studies should be upgraded one level: from low to limited appropriateness and a trustworthiness estimate of 70 percent.

MANY STUDENTS WONDER...

Where do these trustworthiness estimates come from?

As will be discussed in more detail later in this course, we use probability estimates in evidence-based management to describe the trustworthiness (certainty) of a claim, hypothesis, or assumption given the available evidence (e.g., findings from an empirical study). A trustworthiness estimate is quantified as a percentage between 0 and 100, where 100 percent indicates absolute trustworthiness/certainty. Let's imagine that someone claims that companies committing themselves to an ethnically diverse workforce perform better financially. What is the trustworthiness of this claim? When there is no evidence against nor in favor of a claim, trustworthiness is 50/50. It's the same as flipping a coin: the claim could be true or not – we simply don't know because there is no evidence available.

Now, imagine that there is a cross-sectional study that supports this claim. Given that a cross-sectional study is not an appropriate design to demonstrate a cause-and-effect relationship, trustworthiness increases only slightly from 50 percent to 60 percent. This means that there is still a 40 percent chance that a diverse workforce will not yield higher financial performance. However, when there is a randomized controlled trial that supports the claim, the trustworthiness (certainty) would increase to 90 percent.

Of course, this estimation of trustworthiness/certainty also works the other way around. When there is a randomized controlled trial that does not support the claim, the trustworthiness would drop to 10 percent, meaning that there is a 90 percent chance that a diverse workforce will not lead to higher financial performance.

Note that the highest level of trustworthiness is 90 percent, as even the most rigorous and reliable study design is not perfect!

Critical appraisal: Methodological quality

Trustworthiness is also affected by a study's methodological quality, that is, the way it was conducted. Consequently, even a methodologically appropriate study can be untrustworthy. For example, the trustworthiness of a study with a high methodological appropriateness can drop dramatically when the study is not well conducted and contains several weaknesses. In fact, when a study contains too many flaws, its trustworthiness can decline to 50 percent, which means the study has the same predictive value as flipping a coin

When we critically appraise a study's trustworthiness, we therefore start by setting a baseline for its highest level of methodological appropriateness. Based on its number of weaknesses, the level of trustworthiness may then be downgraded by one or more levels. To determine the final level of trustworthiness, you can make use of the following rules of thumb:

- 1 Weakness = no downgrade (we accept that nothing is perfect)
- 2 Weaknesses = downgrade 1 level
- 3 Weaknesses = downgrade 2 levels
- And so on

Learn by doing 7.8

The following excerpts are from an October 2014 article by Nelson Groom, published in the *Daily Mail*, a British newspaper:

Eat Your Greens to Fight the Blues: Scientists Say Eating Large Amounts of Fruit and Vegetables Can Make You Happy!

As the saying goes, "an apple a day keeps the doctor away." Now research suggests that eating large amounts of fruit and vegetables every day can keep the therapist away, too. . . . [Research shows that] eight or more portions [of fruit and vegetables] every 24 hours drastically improves mental well-being. . . . Other results from the study found that fruit has a larger impact on mental health than vegetables, and women experience more of a mental health benefit than men.

The article claims that eating fruits and vegetables (the "cause") has a positive impact on mental health (the "effect"). This claim was based on a cross-sectional study published in 2013 in the academic journal *Social Indicators Research*. Assume the study contained three serious weaknesses.

How would you grade the trustworthiness (certainty) of these findings given the type of question, the study's methodological appropriateness, and its number of weaknesses?

Click here to download the grading table, or use the table at the end of this module

a) High: 90 percent
b) Moderate: 80 percent
c) Limited: 70 percent
d) Low: 60 percent
e) Very low: < 60 percent

Scenario 2: Imagine the study was instead a randomized controlled trial in which a large number of men and women were studied over a 6-month period. In addition, imagine the study contained two serious weaknesses.

Given the type of question, the study's methodological appropriateness, and its number of weaknesses, how would you grade the trustworthiness of its findings?

a) High: 90 percent

b) Moderate: 80 percent

c) Limited: 70 percent

d) Low: 60 percent

Scenario 3: Imagine the study was instead a nonrandomized controlled study that used before and after measurements. In addition, imagine the study contained no weaknesses.

Given the type of question, the study's methodological appropriateness, and its number of weaknesses, how would you grade the trustworthiness of its findings?

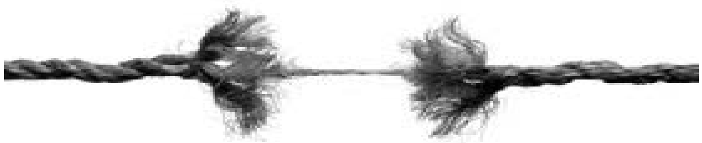
a) High: 90 percent

b) Moderate: 80 percent

c) Limited: 70 percent

d) Low: 60 percent

Methodological quality: Weaknesses (I)



NOTE!

As explained, to determine a study's final level of trustworthiness, we need to assess whether it contains weaknesses. Depending on the study's research design, we ask three to four appraisal questions, discussed next.

Note that it is not necessary to learn these questions by heart. At the end of this module, you can download a checklist with appraisal questions for each study design, and during the assessments and quiz, we will provide you the right questions to ask. For now, just read the following sections to get a general understanding of what questions to ask (and why).

1. Critical appraisal questions to determine weaknesses in a meta-analysis or systematic review

Meta-analyses and systematic reviews summarize studies on the same topic. They often use statistical analysis techniques to combine the results in order to achieve a more precise effect size. To determine their methodological appropriateness, we need to first identify the types of studies included. As the ***grading table*** indicates, a meta-analysis or systematic review based on randomized controlled trials is a very appropriate design to measure an effect, impact, or causal relation; thus, the baseline trustworthiness is very high (95%). When the studies included are cross-sectional (or when it is unclear what type of studies have been included), the baseline trustworthiness drops to 70 percent. To determine the final level of trustworthiness, you can use the following critical appraisal questions:

Q1: Is it likely that important, relevant studies were missed?

As explained in *Module 5, Acquire: Evidence from the Scientific Literature*, the best place to start a comprehensive search for studies is with the major research databases (e.g., ABI/INFORM, Business Source Premier, and PsycINFO). However, a search should also include contact with active researchers, particularly to investigate whether there are any unpublished studies. You can typically find information about relevant studies in the Method section of an article. If the above conditions are met, it is unlikely that key studies would have been missed, so selection bias is minimized.

Q2: Was the process to select studies clearly defined and reproducible?

To prevent selection bias, the choice of which studies to include should preferably be made by at least two reviewers, independently of each other, and using rigorously specified inclusion and exclusion criteria. Furthermore, this process should be clearly documented, for example, in the form of a flowchart that shows how many studies are excluded as a result of which criteria. You can find information about the selection process in the Method section.

Q3: Was the process to extract data clearly defined, and was the outcome presented in a table?

To prevent bias, the extraction of data (such as population, study design, sample size, variables measured, and outcome) should preferably be conducted by at least two reviewers, independently of each other. The outcome should be presented in a table showing clearly the data extracted for each included study. You can find information about the process of data extraction in the Method section, usually under Data Extraction or Coding Procedure.

Q4: Was the methodological quality of each study assessed?

The quality of a systematic review or meta-analysis is determined by the methodological quality of the primary studies included (remember, it's garbage in/garbage out). The quality of the studies included should therefore be appraised using predetermined criteria and validated checklists. This information is typically provided in the Method section. However, keep in mind that a substantial number of systematic reviews or meta-analyses provide insufficient information regarding whether or not the methodological quality of the studies included was critically appraised.

2. Critical appraisal questions to determine weaknesses in a randomized controlled trial

A randomized controlled trial is considered the gold standard to measure an effect, impact, or causal relation. It is not always appropriate for non-effect questions and therefore is seldom used for such questions. A randomized controlled trial, however, can have multiple weaknesses that affect its internal validity.

Q1: Was the control group similar to the intervention group at the start of the study?

If the randomization process worked (i.e., achieved comparable groups), the groups should be similar. As a rule, the more similar the groups are at the start, the better the randomization process is. The study should have a table of baseline characteristics comparing the groups on variables that could affect the outcome. If not, there may be a description of group similarity in the first paragraphs of the Results section. In addition, the e should be an indication of whether differences between groups are statistically significant. If the groups were not similar at the start, we cannot be sure that the outcome of the study is due to the intervention rather than to (unknown) confounding factor(s).

Q2: Did fewer than 20 percent of the subjects drop out?

The percentage of dropouts or withdrawals after the first (baseline) measure should be minimal, preferably less than 20 percent. In most studies, you can find this information in the Results section or the table with the results (e.g., by comparing the number of subjects (*n*) at the start of the study with the number of measurements). If data are collected (or variables measured) from fewer than 80 percent of the original subjects at the start of the study, the outcome may be less trustworthy.

Q3: Were reliable and valid measurement methods used?

As explained earlier, the measurement of direct/objective variables (e.g., production errors, staff turnover rate) is more likely to be valid and reliable than that of self-report/subjective variables (e.g., self-reported accidents or performance). In addition, when a study makes use of a questionnaire, check whether it was developed for the present study ("self-made" or "home-grown") or if it also has been used in other studies. When a scale or questionnaire has been used in previous research, we may assume that it has some track record for validity and reliability, though such information should still be provided in the present study. In most studies, you can find this information in the Method section, usually under Measurements or Tools.

Methodological quality: Weaknesses (II)

3. Critical appraisal questions to determine weaknesses in a nonrandomized controlled before–after study

A nonrandomized controlled study is an appropriate design to measure an effect, impact, or causal relation. Like randomized controlled trials, a nonrandomized controlled design is not appropriate for non-effect questions such as attitudes and frequencies. To determine its methodological quality, we ask some of the same questions that we ask when critically appraising a randomized controlled trial. Additional questions include the following:

Q1. Did the study start prior to the intervention/exposure?

A major drawback of retrospective studies is that they tend to be more susceptible to bias and confounding. Therefore, additional controls – such as blinding – should be applied.

Q2: Was the intervention (or exposure to a variable) independent of other changes over time?

If the intervention or exposure did not occur independently of other changes (e.g., a restructuring program or the implementation of a new management model), or if the outcome was likely to have been influenced by historical events during the study period (e.g., a merger), there may be a negative impact on a study's trustworthiness.

4. Critical appraisal questions to determine weaknesses in a non-randomized controlled study without a pretest

A non-randomized controlled study without a pretest (baseline measure) is not very appropriate for measuring an effect, impact, or causal relationship. Nevertheless, this design is often used when higher-quality methodologies are not possible for practical reasons. To determine its methodological quality, we ask the same questions as when critically appraising a randomized controlled trial.

5. Critical appraisal questions to determine weaknesses in a before–after study or longitudinal study

Before–after studies and longitudinal studies without a control group are not a very appropriate design to measure an effect, impact, or causal relation. For non-effect questions, these type of studies are appropriate to identify antecedents (predictors) or measure differences and associations. To determine their methodological quality, we ask most of the same questions as when critically appraising a controlled study. An additional question is

Q1: Were the criteria used to select subjects clearly defined?

A clear description of the inclusion criteria increases the likelihood that subjects in the study (employees, teams, units, organizations, etc.) were selected because they were representative of the target population rather than for reasons of convenience. In addition, the study should specify how the subjects were selected. Did they invite everyone in the population who met the criteria or just a sample? You can usually find information about the selection criteria in the Method section.

Q2: Were the data gathered for the same subjects repeatedly over a period of time?

A key feature of a longitudinal study is that data (measures) are collected from the *same sample* at different points in time. This means that the study should specify whether the subjects that were measured at the first time point (T1) were the same subjects that were measured at the second and consecutive time points (T2, T3, etc). If it is not clear whether the measurement at the different time points (before and after an intervention or exposure) concerned the same subjects, for instance because an anonymised survey was used, it can't be considered a longitudinal study.

6. Critical appraisal questions to determine weaknesses in a cross-sectional study

Cross-sectional studies such as surveys are appropriate for non-effect questions, especially when they relate to measuring frequencies, opinions, or attitudes. Although they are very often used for studies that aim to measure an effect, impact, or (causal) relation, they are in fact inappropriate for this purpose. We ask the following questions to appraise their methodological quality:

Q1: Was the sample randomly selected?

When a sample is randomly selected, each member of the population has an equal chance of being chosen as a subject in the study. When other methods are used, the study is susceptible to selection bias. The most reliable way to randomly select a sample is to use computer software that generates numbers by chance (such as Excel or SPSS Statistics). You can typically find information about how the sample was selected in the Method section, usually under Sample or Procedure. A randomly selected sample is sometimes called a *probability sample*, whereas a nonrandom sample is called a *convenience sample*.

Q2: Was the sample size large enough?

Whether a cross-sectional study's sample size *(n)* is large enough depends on the population size *(N)*, the margin of error (ME, usually 5%), and the confidence interval (CI, usually 95%). In general, you can use the following rules of thumb as a guideline:

Sample size (ME=5%, P=50%, CI=95%)	
N = 50 > n = 45	N = 500 > n = 218
N = 100 > n = 80	N = 1000 > n = 278
N = 250 > n = 152	N = 10,000 > n = 370

Q3: Is it likely that data dredging has taken place?

Data dredging (also called *data fishing* or *data snooping*) is an inappropriate practice in which a data set is exhaustively analyzed and a large number of hypotheses or relationships tested to find combinations of variables that show a statistically significant correlation. Testing a large number (>20) of correlations or exhaustively analyzing the data, however, increases the chance of detecting a non-existing significant effect (a false effect). You can usually find information about the number of hypotheses or elations (correlations) tested in the Method section (typically under Measurement or Independent/Dependent Variables), or in the table(s) with the results.

7. Critical appraisal questions to determine weaknesses in a qualitative study

As explained, qualitative methods use data obtained from interviews, focus groups, documentary analysis, narrative analysis, or participant observation. This type of method is very appropriate to gain an understanding of underlying reasons, opinions, or motivations of study participants or to generate hypotheses and/or theories that can be tested through subsequent quantitative methodologies. Qualitative methods are not appropriate for measuring an effect, impact, or (causal) relation. To determine the quality of a qualitative method, you can use the following questions: (16)

Q1: Is the researcher's perspective clearly described and taken into account?

In qualitative studies, the researcher is the primary tool for data collection and interpretation. Consequently, in assessing the methodological quality of qualitative studies, the core criterion to be evaluated is researcher bias. To minimize this bias, the researcher should declare his or her assumptions and biases about the topic under study and make a clear statement of how they are likely to have influenced the results. Information about the researcher's assumptions and biases can usually be found in the Research Limitations section, although it might also be found elsewhere. Bear in mind, however, that a substantial number of qualitative studies fail to provide this information.

Q2: Are the methods for collecting data clearly described?

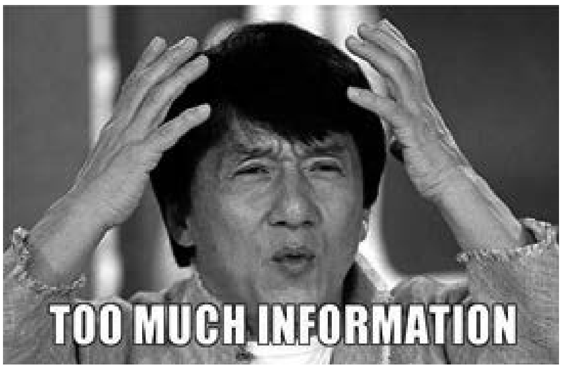
The researcher should provide adequate information about data-collection procedures. For instance, what method was used to collect the data (participant observation, interviews, document reviews, focus groups, etc.)? And in what form did the data come (audio recordings, video material, emails, notes, etc.)? You can find information about how the data we e collected in the Method section, usually under Data Collection or Procedure.

Q3: Are quality-control measures used?

Quality-control measures include the following:

- Independently verifiable data formats (audio recordings, video)
- Independent analysis of data by more than one researcher
- Verbatim quotes, sustained observation over time, peer debriefing (i.e., involving impartial peer researchers to evaluate and make sense of findings
- Addressing negative or discrepant results

You can find information about quality-control measures in the Method section, usually under Data Analysis.



PLEASE NOTE

As explained earlier, it is NOT necessary to learn these questions by heart. At the end of this module, you can download a checklist with appraisal questions for each study design, and during the assessments and quiz, we will provide you the right questions to ask.

Finally: The best available evidence

Let's imagine that we want to know whether the introduction of virtual teams will improve our organization's performance. When we locate a single study in which a meaningful effect was found but a pre-measure was missing, we must conclude, on the basis of what we learned earlier in the article, that this study does not represent the best evidence. It would have been better if the study had used a pre-measure, and it would have been great if a control group had also been used, but it would still not have been the ***best*** evidence. Even a study in which a control group has been randomly assigned cannot be considered the best evidence, as a single study is merely suggestive. Now, as we just have learned, the best study to answer our question would have been a systematic review or meta-analysis of multiple randomized controlled trials with no methodological weaknesses. You probably won't be surprised to hear that such studies are scarce. As a result, the best evidence is often not available. In fact, instead of a meta-analysis of randomized controlled trials, there are often only a handful cross-sectional studies available.

Evidence-based management, however, is not about the best evidence but rather about the best *available* evidence. In the situation described above, the handful of cross-sectional studies constitute the best available scientific evidence. But even this (perhaps disappointing) result could turn out to be important. It indicates that the evidence base on this topic is not (yet) well developed. And that in itself is important information. All managers and leaders have anxieties, and we all question whether what we are doing is right. So, when we find out that the scientific evidence is poor or even absent, it means that the answer to our question is (yet) unknown. And that can be a relief. We are agonizing over which is better, X or Y, but now we know there is (as yet) no definitive answer.

At the same time, it is important to remember that evidence-based management is about the use of evidence from four sources, not just one. That means that regardless of whether the trustworthiness of the scientific evidence you have found is high or low, you should always consider the evidence from multiple sources. No single source of evidence can be viewed as a universal or timeless truth or superior to any other – even a meta-analysis of randomized controlled trials comes with weaknesses and uncertainties. As we stated in the first module of this course, evidence-based practitioners make decisions not based on conclusive, solid, up-to-date information but on probabilities, indications, and tentative conclusions. Scientific evidence rarely tells you what to decide, but it always helps you to make a better-informed decision.

Let's practice! Exercises 1 & 2

Now that you are familiar with the elements that determine a study's trustworthiness, it is time to practice critically appraising scientific studies. Each of the **Learn by doing** activities on this and the next page presents a scientific cause-and-effect study and asks you questions that will help you develop and sharpen your critical appraising skills. Because these four exercises are the first to ask you to critically appraise a full-length published study, most of the questions come with two to three hints to help you out.

Learn by doing 7.9

In July 2013, the following article was originally published on the website of the BBC.

BBC Sign in | News | Sport | Weather | Shop

CAPITAL

Home | Columns | Personal Finance | Careers | DAVOS 2015: Exclusive Coverage >

Funny business: Why humour on the job matters

Eric Barton

Jokes aren't a must for success as a manager, but studies suggest humour leads to creativity, better personal interactions and workplaces that are simply more fun. "Work has an insanely important impact on our lives, so why make it harder than it has to be?" said Michael Kerr, author of You Can't Be Serious! Putting Humour to Work.

Humour can help managers relate with their employees, and that increases the likelihood that a subordinate will believe his boss understands what he does. A well-timed joke in a time of bad news can also make people feel that a manager has things under control. Joking around also correlates with broader keys to success, like accomplishing goals…

The source of the article was a study, "A Meta-analysis of Positive Humor in the Workplace," which was published in the *Journal of Managerial Psychology* in 2012. The results suggest that positive humor in the workplace enhances work performance, satisfaction, and cohesion within work groups, as well as decreases burnout, stress, and work withdrawal.

Download the study here

Determine the study's research design by answering the questions below (yes / no / unclear).

1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Did most included studies use a control group and random assignment?**
3. **Did most of the studies included collect the data (or measure the variables) both before and after exposure to an intervention or variable?**
4. **Based on the answers to the questions above, the study is most likely which of the following?**
   a) A systematic review or meta-analysis based on randomized controlled studies
   b) A systematic review or meta-analysis based on cross-sectional studies
   c) A randomized controlled before–after study
   d) A noncontrolled before–after study
   e) A cross-sectional study

5. How would you consider this design's appropriateness for measuring an effect, impact, or causal relation (in this case, the effect of positive humor on work performance, satisfaction, cohesion, burnout, stress, and work withdrawal)?

Click here to download the grading table or use the table at the end of this module

a) High
b) Moderate
c) Limited
d) Low

Please continue to critically appraise the study's methodological quality.

6. **Is it unlikely that important, relevant studies were missed?**
7. **Was the process used to select studies clearly defined and reproducible?**
8. **Was the methodological quality of each study assessed?**
9. **How large was the effect size? (Use the rules of thumb below as a guideline.)**

Effect size	Small	Medium	Large
Standardized mean difference: d, Δ, g	≤ .20	.50	≥ .80
ANOVA: η^2, ω^2	≤ .01	.06	≥ .14
Chi-square: ω^2	≤ .10	.30	≥ .50
Correlation: r, ρ	≤ .10	.30	≥ .50
Correlation: r^2	≤ .01	.09	≥ .25
Simple regression: β	≤ .10	.30	≥ .50
Multiple regression: β	≤ .20	.50	≥ .80
Multiple regression: R^2	≤ .02	.13	≥ .26

a) Large
b) Medium
c) Small
d) No effect size was reported

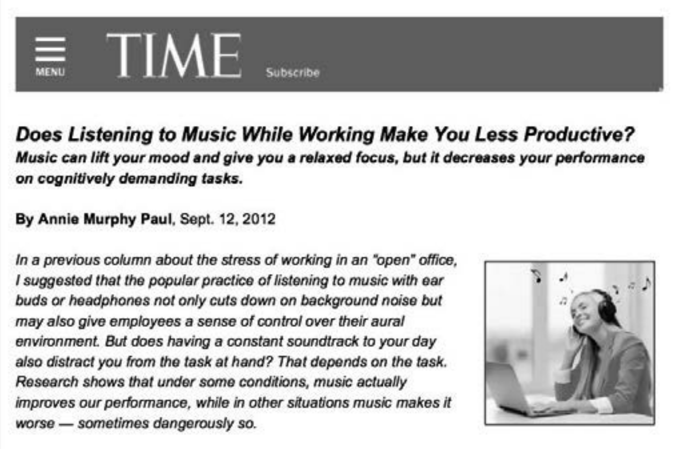
10. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study, and describe the study's methodological quality.**

11. **The findings of the study suggest that there is a positive effect (with a medium effect size of humor on work performance, satisfaction, and work-group cohesion, as well as decreased burnout, stress, and work withdrawal. Given the study's design and methodological quality, how would you qualify the trustworthiness of these findings?**

a) High (90%)
b) Moderate (80%)
c) Limited (70%)
d) Low (60%)

Learn by doing 7.10

In September 2012, the following article was originally published on the website of *Time*:



MENU TIME Subscribe

Does Listening to Music While Working Make You Less Productive?

Music can lift your mood and give you a relaxed focus, but it decreases your performance on cognitively demanding tasks.

By Annie Murphy Paul, Sept. 12, 2012

In a previous column about the stress of working in an "open" office, I suggested that the popular practice of listening to music with ear buds or headphones not only cuts down on background noise but may also give employees a sense of control over their aural environment. But does having a constant soundtrack to your day also distract you from the task at hand? That depends on the task. Research shows that under some conditions, music actually improves our performance, while in other situations music makes it worse — sometimes dangerously so.

The source of the article was a study, “Background Music: Effects on Attention Performance,” that was published in the journal *Work* in 2012. The findings suggest that if background music is played in the work environment, music without lyrics is preferable, because songs with lyrics are likely to reduce worker attention and performance.

Download the study here
and determine the study’s research design by answering the questions below.

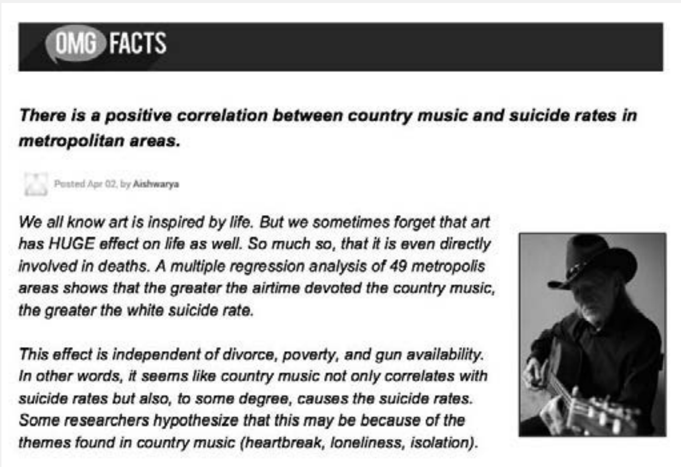
1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Were all the data collected (or variables measured) at the same time?**
3. **Was there a control (comparison) group?**
4. **Were the subjects in the study randomly assigned to an intervention group and a control or comparison group?**
5. **Based on the answers above, how would you describe this study?**
6. **How would you consider this design’s appropriateness for measuring an effect, impact, or causal relation (in this case, the effect of background music on worker attention)?** Use the grading table
   a) High
   b) Moderate
   c) Limited
   d) Low
7. **Was the control group similar to the intervention group at the start of the study?**
8. **Did fewer than 20 percent of the subjects drop out? If not, did the study account for this high number of dropouts?**
9. **Were reliable and valid measurement methods used?**
10. **How large was the effect size of background music with lyrics on attention performance? (Use the rules of thumb** at the end of this module **as a guideline.)**
11. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study, and describe the study’s methodological quality**
12. **The findings suggest that if background music is played in the work environment, music without lyrics is preferable, because music with lyrics is likely to reduce worker attention and performance. Given the study’s design and that there are no serious weaknesses, how would you qualify the trustworthiness of these findings?**
    a) High (90%)
    b) Moderate (80%)
    c) Limited (70%)
    d) Low (60%)

Let’s practice! Exercises 3 & 4

The next two **Learn by doing** activities focus on studies that concern claims about cause and effect.

Learn by doing 7.11

OMG FACTS is a website that provides remarkable facts about a wide range of topics. In April 2002, the following article was published:


OMG FACTS

There is a positive correlation between country music and suicide rates in metropolitan areas.

Posted Apr 02, by Aishwarya

We all know art is inspired by life. But we sometimes forget that art has HUGE effect on life as well. So much so, that it is even directly involved in deaths. A multiple regression analysis of 49 metropolis areas shows that the greater the airtime devoted the country music, the greater the white suicide rate.

This effect is independent of divorce, poverty, and gun availability. In other words, it seems like country music not only correlates with suicide rates but also, to some degree, causes the suicide rates. Some researchers hypothesize that this may be because of the themes found in country music (heartbreak, loneliness, isolation).

The article’s source was a study, “The Effect of Country Music on Suicide,” which was published in the journal *Social Forces* in 1992.

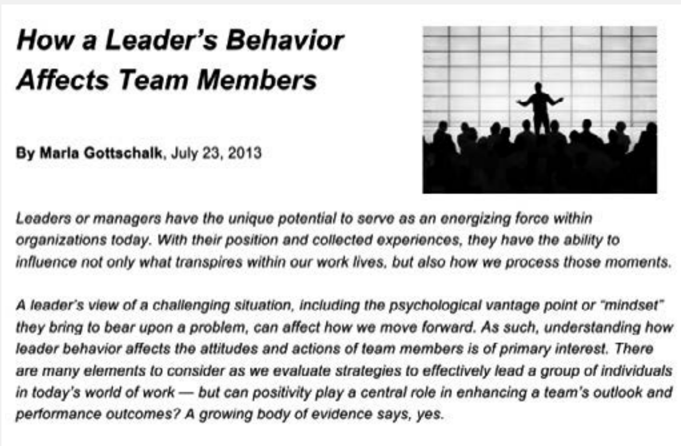
Download the study here
and determine its research design by answering the questions below.

1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Does the study mainly use qualitative data?**
3. **Were all the data collected (or variables measured) at the same time?**
4. **Based on the answers of the questions above, the study is most likely which of the following?**
   a) A systematic review or meta-analysis.
   b) A nonrandomized controlled before–after study
   c) A noncontrolled posttest-only study
   d) A cross-sectional study
5. **Was the sample randomly selected?**
6. **Was the sample large enough?**
7. **Were reliable and valid measurement methods used?**

8. **Is it likely that data dredging has taken place?**
9. **How large was the size of the effect of airtime given to country music and the suicide rate among white people? (Use the rules of thumb as a guideline.)**
10. **How precise was the effect size?**
    a) The 95% CI was OK
    b) The 95% CI was too wide.
    c) There are no CI's reported.
11. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study, and describe the study's methodological quality.**
12. **The findings suggest that listening to country music has a medium to large effect on the suicide rate among white people. Given the study's design and methodological quality, how would you qualify the trustworthiness of these findings?** (use the grading table)
    a) High (90%)
    b) Moderate (80%)
    c) Limited (70%)
    d) Low (60%)
    e) Very low (< 60%)

Learn by doing 7.12

In July 2013, the following article was originally published on the website Switch & Shift.



How a Leader's Behavior Affects Team Members

By Marla Gottschalk, July 23, 2013

Leaders or managers have the unique potential to serve as an energizing force within organizations today. With their position and collected experiences, they have the ability to influence not only what transpires within our work lives, but also how we process those moments.

A leader's view of a challenging situation, including the psychological vantage point or "mindset" they bring to bear upon a problem, can affect how we move forward. As such, understanding how leader behavior affects the attitudes and actions of team members is of primary interest. There are many elements to consider as we evaluate strategies to effectively lead a group of individuals in today's world of work — but can positivity play a central role in enhancing a team's outlook and performance outcomes? A growing body of evidence says, yes.

One of the article's sources was a study, "The Emotional Impact of Leaders' Behaviors," which was published in the *Journal of European Industrial Training* in 2009. The results suggest that a leader who lacks self-management and relationship-management competency induces negative emotions and, as a result, is considered a bad leader by his or her subordinates.

Download the study here and determine the study's research design by answering the questions below.

1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Does the study mainly use qualitative data?**
3. **How would you consider the study design's appropriateness for quantitatively measuring an effect, impact, or causal relation (in this case, the emotional impact of leaders' behavior)?**
    a) High
    b) Moderate
    c) Limited
    d) Low
    e) Not applicable

4. **Is the researcher's perspective clearly described and taken into account?**
5. **Are the methods for collecting data clearly described?**
6. **Are quality-control measures used?**
7. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study, and describe the study's methodological quality.**
8. **The findings suggest that a leader who lacks self-management induces negative emotions and, as a result, is considered a bad leader by his or her subordinates. Given the study's design and the weaknesses found, how would you qualify the trustworthiness of these findings?** (use the grading table)
   a) High (90%)
   b) Moderate (80%)
   c) Limited (70%)
   d) Low (60%)
   e) Very low, like flipping a coin (50%)

Let's practice some more!

This **Learn by doing** exercise concerns a study that is a little more challenging to critically appraise.

Learn by doing 7.13

In June 2014, the following blog was first published on the website of HR & Talent Management, a membership organization for HR managers.

HR&Talent MANAGEMENT

Flextime Implementation for Increased Employee Morale

JUNE 16, 2014 BY VIKKI BEREZOVSKAYA

Good work-life balance is becoming increasingly important to employees, and companies that accommodate for it are becoming increasingly desirable. Employees who have flexibility in how their work week is structured, whether in the form of non-traditional hours, job sharing, or increased telecommuting, report more job satisfaction and less stress. Employers benefit from being accommodating toward their flextime-seeking employees. Content workers are more dedicated, pleasant, and productive. They are tardy and absent from work less frequently than their nine-to-five peers.

One of the blog's sources is a study that was published in the journal *Personnel Psychology* in 1981. This study suggests that workers, when given the opportunity to schedule their own workday, deviate only moderately from their pre-flextime arrival/departure times. In addition, it was found that the number of late arrivals was virtually eliminated

Download the study here
and determine the study's research design by answering the questions below.

1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Were all the data collected (or variables measured) at the same time?**
3. **Was there a control (comparison) group?**
4. **Were all the data collected (or variables measured) both before and after exposure to an intervention or variable?**
5. **Based on the answers above, how would you describe this study?**
6. **How appropriate would you consider this design to be for measuring an effect, impact, or causal relation (in this case, the effect of flexible working hours on the arrival and departure times of employees)?** Use the grading table at the end of this module.
   a) High
   b) Moderate
   c) Limited
   d) Low

7. **Were the criteria used to select subjects clearly defined?**
8. **Was the intervention independent of other changes over time?**
9. **Did fewer than 20 percent of the subjects drop out? If not, did the study account for this high number of dropouts?**
10. **Were reliable and valid measurement methods used?**
11. **How large was the effect size of flextime on the decrease in tardiness (number of late arrivals)? (Use the rules of thumb as a guideline.)**
    a) Large
    b) Medium
    c) Small
    d) No effect
    e) No effect sizes are reported
12. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study, and describe the study's methodological quality.**
13. **The findings of the study suggest that there is a large positive effect of flextime on tardiness. Given the study's design and methodological quality, how would you qualify the trustworthiness of these findings?** (Use the grading table)
    a) High (90%)
    b) Moderate (80%)
    c) Limited (70%)
    d) Low (60%)

Did I get this?

Now you have completed the two **Learn by doing** exercises. It is time to do a quick self-check and assess your own understanding regarding how to evaluate scientific studies. Do you feel that you a e getting it? Use these **Did I get this?** exercises to find out.

The studies in these two exercises concern claims about cause and effect. Because this is a **Did I get this?** exercise, no hints are provided.

Did I get this 7.9

In January 2014, the following article was originally published on the website of *Time* magazine.

MENU TIME Subscribe

What Critics Get Wrong About Financial Education

Doubters of financial education make four great points but reach the wrong conclusion every time.

By Dan Kadlec, Jan. 17, 2014

Is it really possible that dozens of nations, thousands of nonprofits and tens of thousands of educators are wrong to want to teach kids about money in school? That's what critics would have us believe. Certainly there is room for skepticism, and consensus thinking isn't always right after all. Yet crowd thinking usually has a solid foundation.

What started as a given less than two decades ago (if you teach it they will learn) has devolved into high-level handwringing (Where's the proof that financial education works?). ... Recently, author and journalist Helaine Olen has taken up the mantel. ... Olen cites research from 200 studies of financial literacy programs showing that financial education has a negligible impact on subsequent financial decisions and behavior. Within 20 months, almost everyone who had taken a financial literacy class had forgot what they learned.

One of the sources of this article was a study that was published in *Management Science* in 2014. This study shows that the effect of financial education on financial literacy (the degree to which a person understands key financial concepts and has the skills and knowledge to effectively manage his or her personal finances) is minimal.

Download the study here and determine its research design by answering the questions below.

1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Did most included studies use a control group and random assignment?**
3. **How would you consider this design's appropriateness for measuring an effect, impact, or causal relation (in this case, the effect of financial education on financial literacy)?**
   a) High
   b) Moderate
   c) Limited
   d) Low
4. **Is it unlikely that important, relevant studies were missed?**
5. **Was the process used to select studies clearly defined and reproducible?**
6. **Was the process to extract data clearly defined?**

7. **Was the methodological quality of each study assessed?**
8. **How large was the effect size of financial education intervention?** Use the rules of thumb.
   a) Large
   b) Medium
   c) Small
   d) No effect sizes are reported
9. **How precise was the effect size?**
   a) The 95% CI was OK.
   b) The 95% CI was too wide.
   c) There was no CI.
10. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study and describe the study's methodological quality.**
11. **The findings of the study suggest that financial education interventions have a negligibly small effect on financial literacy. Given the study's design and the three weaknesses found, how would you qualify the trustworthiness of these findings?** Use the grading table.
   a) High (90%)
   b) Moderate (80%)
   c) Limited (70%)
   d) Low (60%)

Did I get this 7.10

In 2009, *Harvard Business Review* published an article on coaching.

HBR Research Report

BY DIANE COUTU AND CAROL KAUFFMAN

What Can Coaches Do for You?

The coaching field is filled with contradictions. Coaches themselves disagree over why they're hired, what they do, and how to measure success. Here's what you should know.

IN THE SEVENTEENTH CENTURY, the French statesman Cardinal Richelieu relied heavily on the advice of Father François Leclerc du Tremblay, known as France's *éminence grise* for his gray monk's habit. Like the famous cardinal, today's business leaders have their gray eminences. But these advisers aren't monks bound by a vow of poverty. They're usually called executive coaches, and they can earn up to $3,500 an hour.

To understand what they do to merit that money, HBR conducted a survey of 140 leading coaches and invited five experts to comment on the findings. As you'll see, the commentators have

The main point of this piece was that coaching as a business tool continues to gain legitimacy, but there is not much empirical evidence to support that coaching works for everybody in all circumstances. The authors therefore conclude that when it comes to coaching, it is with the caveat "buyer beware!" A year later, "Enhancing Performance: A Case Study of the Effects of Employee Coaching in Construction Practice" was published in *Construction Management and Economics*. The study suggested that coaching significantly enhances employees' personal and professional growth, motivation level, and performance.

Download the study here
determine the study's research design by answering the questions below.

1. **Does the study mainly examine the results of a large number of studies published on the same topic?**
2. **Does the study mainly use qualitative data?**
3. **Were all the data collected (or variables measured) at the same time?**
4. **Based on the answers above, how would you describe this study?**
5. **How would you consider this design's appropriateness for measuring an effect, impact, or causal relation (in this case, the effect of coaching on employee performance)?** Use the grading table at the end of this module.
   a) High
   b) Moderate
   c) Limited
   d) Low
6. **Was the sample randomly selected?**
7. **Was the sample large enough? (Use the rules of thumb as a guideline)**
8. **Were reliable and valid measurement methods used?**
9. **Is it unlikely that data dredging has taken place?**
10. **How large was the effect size of coaching on employees' organizational skills? (Use the rules of thumb as a guideline.)**
    a) Large
    b) Medium
    c) Small
    d) No effect
    e) No effect sizes are reported
11. **Briefly summarize the findings and design of this study. Identify any serious weaknesses in the study, and describe the study's methodological quality.**
12. **The findings suggest that coaching significantly enhances employees' personal and professional growth, motivation level, and performance. Given the study's design and the three serious weaknesses found, how would you qualify the trustworthiness of these findings?** Use the grading table.
    a) High (90%)
    b) Moderate (80%)
    c) Limited (70%)
    d) Low (60%)
    e) Very low (< 60%)

- If we find a study that appears to answer our question, we need to know the trustworthiness of that answer How certain can we be that the findings found in the study a e valid and reliable? Or put differently, what is the chance that alternative explanations for the observed effect are possible?
- The trustworthiness of a study is first determined by its methodological appropriateness: Is the way the study is designed the best way to answer the research question? A randomized controlled study, for instance, is considered to be a better (more appropriate) design to answer questions about cause and effect than in a cross-sectional study.
- A study's trustworthiness is also determined by its methodological quality – the way the study was conducted. Based on its number of weaknesses, the trustworthiness of a study may be downgraded by one or two levels. Thus, a randomized controlled study with several serious flaws can be less trustworthy than well-conducted nonrandomized controlled study with no weaknesses.
- Relevant questions to determine a study's methodological quality are, for example,
    1. Was the sample size large enough?
    2. Were the inclusion criteria used to select subjects clearly defined
    3. Were reliable measurement methods used?
- When critically appraising a scientific study, it is also important to look at the effect size. An effect (or difference) can be statistically significant, but it is not necessarily of practical relevance.
- When you have determined a study's methodological appropriateness and quality, you can determine its final level of trustworthiness. A level of trustworthiness of 70% (limited) means there is a 30% or more chance that alternative explanations for the effect (or difference) found are possible. A level of trustworthiness of 50% (zero) means that the study has the same predictive value as flipping a coin

Podcast: APPRAISE Evidence from the Scientific Literature

In this podcast host Karen Plum discusses with Eric Barends, Managing Director of CEBMa, Denise Rousseau, Professor at Carnegie Mellon University, and Rob Briner, Professor of Organizational Psychology at Queen Mary University of London how to evaluate the quality of the evidence from the scientific literature that we've identified during our earlier searching.




Modules 5, 6 and 7 all focus on the scientific literature, so when you listen to the corresponding podcast episodes, the picture will hopefully be complete.

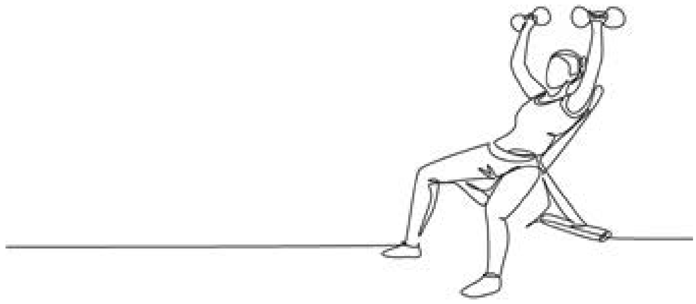
In this podcast we look at the process of learning how to appraise the studies, and what their findings mean in the context of the real world that we are faced with. If something is statistically significant and the e is a large effect size measured by the study, does that mean that the finding is practically relevant for our organisation?

Evidence from the scientific literature is undoubtedly important, but shouldn't be emphasized more than other sources of evidence and it's good to be reminded that it's only when these come together than we can really make an evidence-based decision.

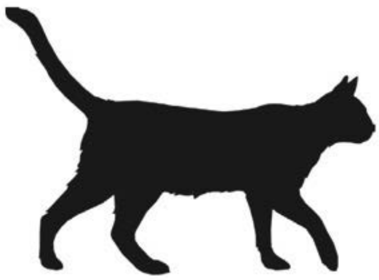
Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 7.1: Critically Appraised Topic (CAT)



Conduct a CAT on a topic relevant to your organization/project, or that sparks your professional interest. Present your findings in a Word document according to ***CEBMa's CAT Guideline***

More detailed instructions will be provided by your professor/instructor along with examples of CATs.

TIPS:

1. Before you start working on your CAT report, make sure that your CAT question is approved by your professor/instructor (we want you to do a CAT on a topic that guarantees learning and for which sufficient scientific evidence is available
2. Before you start critically appraising and summarizing the studies, ask your professor/lecturer to check your search strategy.
3. Before you submit your CAT report, use the checklist in the CAT Guideline to make sure you have covered all CAT elements.

Assessment

Your CAT will be assessed according to the criteria listed in the ***CAT-assessment form***. You will receive feedback and a preliminary grade for your first draft if turned in by the due date. You then have about a week to reflect on the feedback, make changes if needed, and submit your final CAT report. Your instructor will review the changes you made and then determine your final grade.

Suggestions for further reading

- *It's the Effect Size, Stupid. What effect size is and why it is important*, Robert Coe, 2002
- *Effect sizes and the interpretation of research results in international business*, Ellis, JIBS, 2010
- *Not Even Scientists Can Easily Explain P-values*, Christie Aschwanden, FiveThirtyEight Science Blog, Nov 2015
- *Evidence, hierarchies, and typologies: horses for courses*, Petticrew & Roberts, JECH, 2003
- *GRADE: an emerging consensus on rating quality of evidence and strength of recommendations*, Guyatt et all, BMJ, 2008
- *Difficult but Doable: Increasing the Internal Validity of Organizational Change Management Studies*, Barends et al, JABS, 2014
- *Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research*, Sarah Tracy, Qualitative Inquiry, 201

Cohen's Rules of Thumb

According to Cohen, a small effect is one that is visible only 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. The following table provides an overview of how this rule of thumb applies to different measures of effect sizes.

Effect size	***Small***	***Medium***	***Large***
Independent means: d, Δ, g	≤.20	.50	≥.80
Correlation: r, ρ	≤.10	.30	≥.50
Correlation: r^2	≤.01	.09	≥.25
ANOVA: f	≤.10	.25	≥.40
ANOVA: η^2	≤.01	.06	≥.14
Simple regression: β	≤.10	.30	≥.50
Multiple regression: β	≤.20	.50	≥.80
Multiple regression: R^2	≤.02	.13	≥.26
Multiple regression: f^2	≤.02	.15	≥.35

Note, however, that Cohen’s rules of thumb were meant to be exactly that – rules of thumb – and are for many reasons arbitrary. For example, a d 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, we need to relate the effect size directly to the outcome measured and its relevance, importance, and meaning in each specific context.

Grading Table

Methodological Appropriateness (Research questions about Cause and Effect)		
Research design	**Appropriateness**	**Trustworthiness**
• Meta-analysis or systematic review of randomized controlled trials	High	> 90%
• Randomized controlled trial • Meta-analysis or systematic review of nonrandomized controlled before–after studies		90%
• Nonrandomized controlled before–after study • Meta-analysis or systematic review of controlled studies without a pretest • Meta-analysis or systematic review of before–after studies without a control group	Moderate	80%
• Controlled study without a pretest • Before–after study without a control group • Longitudinal study without a control group • Meta-analysis or systematic review of cross-sectional studies	Limited	70%
• Cross-sectional study	Low	60%
• Qualitative study, case study	not applicable	50/50

Methodological Quality Based on the number of methodological weaknesses, downgrade the level of trustworthiness by one or more levels	
1 Weakness	No downgrade (we accept nothing is perfect)
2 Weaknesses	Downgrade 1 level
3 Weaknesses	Downgrade 2 levels
4 Weaknesses	Downgrade 3 levels

References

1. Collins, J. C. (2001). *Good to great: Why some companies make the leap . . . and others don't*. Random House. London.

2. Field, A. (2012). *Discovering statistics using IBM SPSS statistics.* Los Angeles: Sage.

3. Ellis, P. D. (2010). T*he essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge: Cambridge University Press.

4. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.

5. Cooper, H. M., & Lindsay, J. L. L. (1998). *Research synthesis and meta-analysis.* Los Angeles: Sage.

6. Note that direct or objective measures can also be unreliable, so it is important to examine how such measures were obtained and whether they were checked for reliability and validity.

7. Lazarsfeld, P. F., & Rosenberg, M. (1955). *The language of social research: A reader in the methodology of social research.* Free Press. New York.

8. Salkind, N. J. (2016). *Statistics for people who (think they) hate statistics*. Sage Publications. Los Angeles.

9. More hilarious examples of spurious correlations can be found at http://www.tylervigen.com/spurious-correlations.

10. Technically this is not completely accurate, as there could be a so-called suppressor effect due to a third (unknown) variable. For example, suppose that a researcher is interested in the relationship between workers' intelligence and the number of errors made on an assembly line task. It seems plausible that, the more intelligent workers would make fewer errors, so we would expect a strong negative correlation. However, the more intelligent workers could also exhibit higher levels of boredom, and boredom would be correlated with the number of errors. In that case, the effect of intelligence on errors would be negative, and the effect of boredom on errors would be positive. Combined, these two effects may cancel each other out, resulting in a total correlation of intelligence on errors equal to zero.

11. The example above (note 10) is adapted from MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention science*, 1(4), 173-181.

12. Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). "Subjective well-being and adaptation to life events: A meta-analysis." *Journal of Personality and Social Psychology*. Volume 102. Number 3.

13. Again, technically this is not completely accurate. When a study uses random assignment, all potential confounding variables that may affect the outcome are equally distributed among the groups. As a result, they are similar at baseline, so a pretest is not strictly necessary. However, without a pretest, we don't know whether the randomization was successful and thus we can't be certain that the groups were truly similar at baseline. For this reason, even in a randomized controlled trial, a pretest is often used.

14. Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). "). Surveying the forest: A meta, Äêanalysis, moderator investigation, and future, Äêoriented discussion of the antecedents of voluntary employee turnover." *Personnel Psychology*. Volume 71. Number 1. 23-65 Pages.

15. Wu, C. Y., Lee, H. F., & Liu, H. (2020). "Effect of temperature and precipitation change on crime in the metropolitan area in Virginia, USA." *Asian Geographer.* Volume 37. Number 1. 17-31 Pages.

16. Tracy, S. J. (2010). "Qualitative quality: Eight 'big-tent' criteria for excellent qualitative research." *Qualitative Inquiry*. Volume 16. Number 10. 837-851.

Learn by doing & Did I get this? Answers and Feedback

Learn By Doing 7.1

a: Correct

b: Incorrect

c: Correct

d: Incorrect

Learn by Doing 7.2

a: Correct

b,c: Incorrect > An η^2 of .065 and a *d* of .59 are both considered a medium difference.

d: Incorrect > The good thing about effect sizes and Cohen's rules of thumb is that they are comparable.

Did I Get This 7.1

a: Correct > According to Cohen's rules of thumb, a β of .30 and .33 are considered a medium effect.

b: Incorrect

c: Correct > The difference between .30 and .33 is not relevant from a practical perspective.

d: Incorrect

e: Incorrect

Learn By Doing 7.3

a: Correct > Almost all confidence intervals have a value at the lower end that is negative and a value at the upper end that is positive. This indicates that there is a 95 percent certainty that the true effect size could be negative or positive. Obviously, if the effect on performance were negative, we would make a different decision than if the effect were positive; thus, the confidence intervals a e too wide.

b: Incorrect > see above

c: Incorrect > There is no such thing as a confidence interval that is too small. After all, an ext emely small 95% CI would mean that there is a 95 percent confidence that e fect size found is the same as the true effect size.

d: Incorrect

Did I get This 7.2

a: Correct > The confidence interval indicates that the e is a 95 percent certainty that the true effect size will be between η^2 = .18 and η^2 = .27, which are both considered a large effect. Thus, the confidence interval is sufficiently nar ow.

b,c,d: Incorrect

Did I Get This 7.3

1: D

2: A

3: C

4: B

Did I Get This 7.4

1: D

2: B

3: A

4: C

Did I get This 7.5

1: Non-effect

2: Cause-and-effect

3: Cause-and-effect

4: Non-effect

5: Cause-and-effect

Learn By Doing 7.4

Study 1

a: Incorrect

b: Correct

c: Correct

d: Incorrect

e: Incorrect

Study 2

a: Correct > If there is no correlation, there is most likely no causal relation.

b: Incorrect > The findings a e rather clear: no correlation was found.

c: Incorrect > The findings a e rather clear: no correlation was found. If there is no correlation, there is most likely no causal relation.

d: Almost right > You are correct that the difference between the impact of highly job-related diversity and less job-related diversity on performance is zero (after all, neither type of diversity had a relationship with performance), but there is better conclusion.

Learn By Doing 7.5

a: Incorrect > The sales agents were not randomly assigned to different groups, so not all criteria for causation are met.

b: Correct > The study measured the effect of two variables (self-set and assigned goals) on sales performance, used two different groups (agents with self-set goals and agents with assigned goals), and used a pre- and posttest measurements (sales data were collected both before and after the goals were set). Thus, three out of four criteria were met.

c: Incorrect > see above

d: Incorrect > see above

Did I get This 7.6

a,b,c: Incorrect > This is a cross-sectional study, meaning that all data were collected at the same point in time. Although a comparison is made between salespersons with a customer orientation and sales agents with a sales orientation, neither group can be considered a (randomized) control group.

d: Correct

Did I get This 7.7

a: Correct > This study is a randomized controlled trial, the gold standard to answer cause-and-effect questions.

b,c,d: Incorrect

Learn By Doing 7.6

a: Almost right > You are correct that the researchers want to know how people feel about the safety of their neighbourhood, but their main question is whether perception of safety is related with a person's weight.

b: Incorrect > The researchers don't want to know how many people with a high BMI perceive their neighborhood as less safe; they only want to know whether these two variables are related.

c: Incorrect > The researchers don't want to know whether a high BMI *causes* a person to perceive his or her neighborhood as less safe; they only want to know whether these two variables are related.

d: Correct > The researchers want to know whether there is an association between a person's weight and how he or she perceives the safety of the neighborhood.

Learn By Doing 7.7

a: Incorrect

b: Incorrect

c: Incorrect

d: Correct

Did I get This 7.8

a: Incorrect

b: Incorrect

c: Correct

d: Incorrect

Learn By Doing 7.8

Scenario 1

a,b,c,d: Incorrect

e: Correct > A cross-sectional study lacks a control group and a pretest and is therefore not appropriate to answer questions about cause and effect. The baseline regarding the study's trustworthiness is therefore low (60%). However, this study also contains three serious weaknesses that further downgrade its trustworthiness. Consequently, we must reduce the study's trustworthiness two levels. This means there is a more than 40 percent chance that alternative explanations (other than eating large quantities of fruits and vegetables) are possible for the improvement of mental well-being.

Scenario 2

a,c,d: Incorrect

b: Correct > The study is a randomized controlled study, a design that uses a control group and random assignment. A randomized controlled study is therefore appropriate to answer questions about cause and effect, so its baseline regarding the study's trustworthiness is high (90%). However, the study contains two serious weaknesses that negatively affect its trustworthiness. Consequently, we must downgrade the study's trustworthiness by one levels, to moderate (80%). This means there is a 20 percent chance that alternative explanations for the improvement in mental well-being are possible.

Scenario 3

a,c,d: Incorrect

b: Correct > A nonrandomized controlled study that uses before and after measurements lacks random assignment and is therefore moderately appropriate to answer questions about cause and effect. The baseline regarding the study's trustworthiness is therefore moderate (80%). In addition, the study did not contain any weaknesses, so the trustworthiness remains 80 percent. This means there is only a 20 percent chance that alternative explanations for the improvement of mental well-being are possible.

Learn By Doing 7.9

1: Yes > Correct.

No > Incorrect

When you look at the abstract, you can read under Design/Methodology/Approach, "The authors examine the results of prior research." This means the study does not examine original, firsthand data collected by the researchers themselves (primary research) but instead examines the results of previously published research (secondary research). The study, therefore, is a systematic review or meta-analysis.

2: Yes > Incorrect

No/unclear > Correct

When you look at the abstract, you can read under Design/Methodology/Approach, "The authors examine the results of prior research." This means the study does not examine original, firsthand data collected by the researchers themselves (primary research) but instead examines the results of previously published research (secondary research). The study, therefore, is a systematic review or meta-analysis.

3: Yes > Incorrect

No/unclear > Correct

The Method section does not provide information regarding whether the included studies collected the data both before and after exposure to an intervention or variable. We must therefore assume that all included studies collected the data at the same time and were most likely cross-sectional studies or surveys, which makes the study more prone to bias.

4. a: Partly correct

b: Correct

c: Incorrect

d: Incorrect

e: Incorrect

This study examines the results of previously published research, which makes this study most likely a systematic review or meta-analysis. However, no information is provided regarding the included studies, so we must assume that they are most likely cross-sectional in nature.

5. a: Incorrect

b: Reasonable

c: Correct

d: Reasonable

The study design is a systematic review or meta-analysis based on cross-sectional studies, a design that lacks a control group, random assignment, and before and after measurements. A cross-sectional study is therefore of low appropriateness to measure an effect, impact, or outcome. However, because this is a systematic review or meta-analysis, we must upgrade its appropriateness by one level, to limited.

6. Yes: Correct

No/unclear: Incorrect

The Method section (first paragraph) states that a computerized sea ch of the PsycINFO and ABI Inform databases was conducted using relevant keywords or phrases. In addition, a manual search of references cited in recently published reviews was conducted. Although the authors did not contact experts to inquire about unpublished research, it is unlikely that important, relevant studies were missed.

7. Yes: Incorrect

No/unclear: Correct

The Method section (under Database, paragraph 2) provides only one inclusion criterion: a study must have reported a relationship between supervisor/employee humor and at least one relevant correlate. No information is provided about how the selection process was conducted (were the studies selected by two independent reviewers?) or how many studies were excluded and why.

8. Yes: Incorrect

No/unclear: Correct

The Method section provides no information about the critical evaluation of the included studies' methodological quality.

9. a: Not quite right

b: Correct

c: Not quite right

d: Incorrect

As you can read in the first paragraph of the Results section, several e fect sizes are provided. However, the effect sizes, in this case rho/ρ, are all in the range of 0.16 to 0.45, which can be considered a medium effect.

10. In this study, the effect of positive humor in the workplace on employee health and work-related outcomes (e.g., performance, job satisfaction, withdrawal) is examined. The study design is a systematic review or meta-analysis based on cross-sectional studies. The methodological appropriateness of this design to measure an effect, impact, or causal relation (in this case, the effect of positive humor in the workplace on employee health and work-related outcomes) is *limited*. In addition, the study has *two serious weaknesses*: (1) The process used to select studies was not clearly defined and is not eproducible. (2) The methodological quality of each included study was not assessed.

11. a: Incorrect

    b: Not quite right

    c: Reasonable

    d: Correct

    The study design is limited appropriate to measure an effect (= 70%). In addition, the study contains two serious weaknesses that affect its trustworthiness. Consequently, we must downgrade the study's trustworthiness by one level, to low (60%). In terms of probability, we could say that the chance that humor has a (medium) effect on employee health and work-related outcomes is about 60%. This means that there is a 40% chance that there is no effect or that the effect is caused by factors other than humor.

Learn By Doing 7.10

1. Yes > Incorrect

   No > Correct

   When you look at the abstract, you can read that 102 participants were recruited to examine the effect of background music on their concentration and attention. This means the study does not examine the results of previously published research (secondary research) but instead examines original, firsthand data collected by the esearchers themselves (primary research).

2. Yes > Incorrect

   No > Correct

   In the Methodology section under 2.3 Procedure, you can see that the data were collected when the participants were in a quiet environment (step 2) and when there was background music (step 3). This means the data were collected both before and after the intervention (background music with lyrics vs. background music without lyrics). An overview of both before and after measures is given in Table 2.

3. Yes > Correct

   No > Incorrect

   In the Methodology section, you can see under 2.3 Procedure that the 102 participants were divided into two groups. Group 1 listened to music with lyrics, and Group 2 listened to music without lyrics. You could have been misled by the fact that both groups were exposed to the intervention (background music), so there isn't a true experimental and control group. However, the fact that a comparison was made between the effect of background music with lyrics and background music without lyrics makes this a controlled study.

4. Yes > Correct

   No > Incorrect

   In the Methodology section under 2.3 Procedure, it states that the 102 participants were randomly divided into two groups.

5. The study is a quantitative study that uses a control group, random assignment, and before and after measurements. This makes the study a randomized, controlled study.
6. a: Correct

   b: Not quite right

   c: Incorrect

   d: Incorrect

   A randomized, controlled study is a design that uses a control group, random assignment, and before and after measurements, and is therefore appropriate for measuring an effect, impact, or outcome. However, in contrast to a systematic review or meta-analysis, it is based on only one study, so the appropriateness of a single randomized, controlled study is regarded as high.
7. Yes: Correct

   No: Incorrect

   In this study the 102 participants were randomly divided into two groups, so we may assume that the groups were similar at the start of the study. You can also read this in the first paragraph of th Results section. An overview of baseline measures is provided in Table 1.
8. Yes: Correct

   No: Incorrect

   The Results section does not state that some participants dropped out. In addition, you can see in Table 2 that the combined number of participants in the two groups (49 + 53) is exactly the same as the number of participants at the start of the study (102).
9. Yes: Correct

   No: Incorrect

   In the Methodology section, you can read under 2.2 Tools that the researchers used the Chu Attention Test, a standard tool for evaluating occupational conditions in Chinese-speaking populations. The researchers also provide a reference for this test, so we may assume that this test is valid and reliable.
10. Large: Incorrect

    Medium: Incorrect

    Small: Incorrect

    No effect size was reported: Correct

    As you can read in the Results section and see in Table 2, no effect sizes are given, only an average score (mean) and a standard deviation (SD). This means the study demonstrates that background music has a negative effect on attention and concentration, but it fails to explain how large this effect is.
11. The study design is a randomized, controlled study. This design is appropriate for measuring an effect, impact, or causal relation (in this case, the effect of background music on worker attention). In addition, the study has no serious weakness. The study, however, does not report effect sizes.
12. a: Correct

    b: Reasonable

    c: Incorrect

    d: Incorrect

    The study design is highly appropriate for measuring an effect, and has no serious weaknesses. We should therefore conclude that the trustworthiness of the findings is high. In terms of p obability, we could say that the chance that background music affects worker attention is about 90%. This means that there is a 10% chance that alternative explanations for the effect found are possible.

    However, note that the study fails to provide any information regarding the size of the effect, so from a practical perspective, the study's usefulness is limited.

Learn By Doing 7.11

1. Yes: Incorrect

   No: Correct

   In the Methodology section, you can read that the researchers used data on music and suicide rates from 49 metropolitan areas. This means the study does not examine the results of previously published research (secondary research) but instead examines original, firsthand data collected by th researchers themselves (primary research).

2. Yes: Incorrect

   No: Correct

   In the Methodology section, you can read in the first paragraph that exposu e to country music was measured as the proportion of radio airtime devoted to country music. In addition, in the second paragraph, you can read that the suicide data were expressed in ”the number of suicides per 100,000 population.” Proportions and numbers cannot be considered as qualitative data.

3. Yes: Correct

   No: Incorrect

   In the first paragraph of the Methodology section, you can ead that the researchers took a single sample. The data in this sample regarding exposure to country music are from one period only, Spring 1985. The data regarding suicide are an average of the 1984 and 1985 rates (see second paragraph). This means there were no before and after measurements, and all data were collected in a single time period.

4. a: Incorrect > This study does not examine the results of previously published research, so it is not a systematic review or meta-analysis.

   b: Incorrect > This study does not include a control group and did not use before and after measurements, so it is not a controlled before–after study.

   c: Incorrect >In this study, all data were collected at the same time, so it is not a noncontrolled posttest-only study.

   d: Correct >In this study, all data were collected at the same time, so it is a cross-sectional study.

5. Yes: Incorrect

   No: Correct

   The first sentence of the Methodology section states: “Our sample is comprised of 49 larg metropolitan areas for which data on music were available.” This means that the researchers did not use a random sample but a sample containing data from specific met opolitan areas. As there are more than 300 metropolitan areas in the United States and there may be a reason why some of metropolitan areas collect data on music and some don’t, the sample the researchers used is susceptible to selection bias.

6. Yes: Incorrect

   No: Correct

   There are 381 metropolitan areas and the researchers collected data from only 49. With a total number of 381 the sample size should be at least 150.

7. Yes: Correct

   No: Incorrect

   In the Methodology section, you can read that the proportions of radio airtime devoted to country music were obtained from the *Radio and Records Rating Report and Directory*. The annual mortality rates were obtained from the Inter-University Consortium of Political and Social Research (ICPSR) at the University of Michigan. Although the study does not provide information on how exactly these proportions and rates were calculated, we may assume that the data are reliable.

8. Yes: Incorrect

   No: Correct

   The study researches only the effect of country music on suicide rates; no other research questions or hypotheses are mentioned. The study does also measure the effect of four other variables (structural poverty, southern region, divorce, and gun availability), but this cannot be considered as data dredging.

9. Large: Incorrect

   Medium: Correct

   Small: Incorrect

   No effect size was reported: Incorrect

   As you can read in the first sentence of the Results section, the cor elation between the suicide rate of white people and airtime given to country music is r = .54, which is considered a large effect. However, because the researchers suspected that other variables, such as structural poverty, living in a southern region, being divorced, and gun availability, could also have an effect on the suicide rate, they conducted a multiple regression to determine the contribution of each variable. The results are reported in Table 1. Under Panel A you can see that the contribution of airtime given to music is β = .27, which is considered a small to medium effect. As you can see, all five variables together have a effect of R^2 = .52, which is considered a large effect. This means that these five variables account fo 52% of the increase (or decrease) of the suicide rate.

10. a, b: Incorrect

    c: Correct

    As you can see in the Method section and in Table 1, no confidence intervals we e provided.

11. In this study, the researchers examine whether country music has an effect on metropolitan suicide rates. The results suggest that the greater the radio airtime devoted to country music, the higher the suicide rate among white people. The study design is a cross-sectional study. This design is not appropriate for measuring an effect, impact, or causal relation (in this case, the effect of country music on suicide rates). In addition, the study has two serious weaknesses: (1) The sample was not randomly selected. (2) The sample was not large enough.

12. a,b,c,d: Incorrect

    e: Correct

    The study design is not appropriate for measuring an effect, and the study has two serious flaws which makes it very prone to bias. We should therefore conclude that the trustworthiness of the findings is (very) lo . In terms of probability, we could say that the chance that country music has an effect on the suicide rate among white people is less than 60%. This means that there is a 40% ot higher chance that there is no effect or that the effect is caused by factors other than country music.

Learn By Doing 7.12

1. Yes: Incorrect

   No: Correct

   When you look at the first two paragraphs of the Methodology section, you can ead that the researchers asked 42 MBA students how the behavior of managers/leaders created emotional responses. This means that the study does not examine the results of previously published research (secondary research) but instead examines original, firsthand data collected by the esearchers themselves (primary research). You may have been misled by the fact that in the sections preceding the Methodology section, the researchers refer to a lot of other studies on the same or related topics. However, the various outcomes of these studies are not systematically combined to arrive at an evaluative, overall conclusion.

2. Yes: Correct

   No: Incorrect

   In the Methodology section, you can read under Fieldwork and Data-Collection Instrument that participants were asked the following question: “Think about the best leader and the worst leader you have ever had. What behaviors did they show (i.e., what did they do?) that made them so great or so bad?” This is an open question to which the answer cannot be expressed in numbers. This in contrast to closed questions to which only a limited number of answers are possible, which makes it possible to quantify the outcome.

3. a,b,c: Incorrect

   d: Reasonable

   e: Correct

   A qualitative study is a design that does not quantitatively measure an outcome. Its appropriateness for measuring an effect, impact, or outcome is therefore (very) low, or better: not applicable.

4. Yes: Incorrect

   No: Correct

   Although it is mentioned twice that interpretation of the responses was based on the researchers’ knowledge of a certain model (Goleman’s model), which could be considered to be subjective (e.g., in the abstract under “Research limitations/implications” and in the first paragraph under “Findings”), th researchers fail to provide information about their preexisting assumptions and biases.

5. Yes: Correct

   No: Incorrect

   In the Methodology section, under Fieldwork and Data-Collection Instrument, you can see that the students were asked a question and then wrote down their answers. In addition, under Analytical Approach, the authors state that students were asked to read their responses out loud and were then asked to write down their feelings.

6. Yes: Incorrect

   No: Correct

   Quality-control measures are mentioned in neither the Methodology section nor in any of the other sections.

7. In this study, the researchers examine the emotional impact of leaders’ behavior. The results suggest that a leader who lacks self-management and relationship-management competency induces negative emotions and, as a result, is considered a bad leader by his or her subordinates. The study design is a qualitative study, which is not appropriate for quantitatively measuring an effect, impact, or causal relation (in this case, the emotional impact of leaders’ behavior). In addition, the study has two serious weaknesses: (1) The researchers fail to provide information about their preexisting assumptions and biases. (2) No quality-control measures were used.

8. a,b,c: Incorrect

   d: Reasonable

   e: Correct

   The study design is not appropriate for measuring an effect. In terms of probability, we could say that the chance that a leader who lacks self management is perceived as a bad leader is therefore about 50/50 (like flipping a coin)

Learn By Doing 7.13

1. Yes: Incorrect

   No: Correct

When you read the first paragraph of the Method section, you can see that the study examines data collected in four government agencies in Israel. It is therefore primary research: a study that examines original, firsthand data collected by the esearcher(s). This is in contrast to secondary research, such as a systematic review or meta-analysis, which examines the results of previously published research.

2. Yes: Incorrect

   No: Correct

   When you read the first paragraph of the Method section, you can see that the data we e collected six months prior to and six months after the implementation of the flextime p ogram. An overview of both before and after measures is given in Table 1.

3. Yes: Incorrect

   No/unclear: Correct

   When you read the first paragraph of the Method section, you can see that a g oup of 162 employees was exposed to the intervention (i.e., given flexible working hours), but the e was no control group of employees who were not exposed to it (i.e., kept traditional, fixed working hours)

4. Yes: Correct

   No/unclear: Incorrect

   When you read the first paragraph of the Method section, you can see that the data we e collected six months before and six months after the implementation of the flextime p ogram. An overview of both before and after measures is given in Table 1.

5. The study is a quantitative study with before and after measurements but without a control group. This study is therefore a non-controlled before–after study.

6. a: Incorrect

   b: Not quite right

   c: Correct

   d: Not quite right

   A non-controlled before–after study lacks a (randomized) control group, which makes the study prone to bias and confounding. The appropriateness of the study for measuring an outcome of effect is therefore limited.

7. Yes: Incorrect

   No/unclear: Correct

   Although the Method section states that the employees were randomly selected, it fails to provide a clear description of the inclusion criteria. It states only that 162 employees "*fulfilled the research requirements*."

8. Yes: Correct

   No: Incorrect

   The study does not state that other changes occurred, so we may assume that this was not the case.

9. Yes: Correct

   No: Incorrect

   The Method section states that 162 employees fulfilled the esearch requirements, and the Results section states that data available for 160 workers were compared before and after the implementation of flextime. This means that only two employees d opped out, which is less than 20 percent.

10. Yes: Correct

    No: Incorrect

    The first paragraph of the Method section states, "Arrival and departu e times were gathered from individual time-clock records." This can be considered to be a valid and reliable measurement method.

11. a: Reasonable

    b,c,d: Incorrect

    e: Correct

    The Results section states that prior to the implementation of flextime ta diness was excessive, but after its implementation, tardiness was drastically reduced. The number of late arrivals decreased from an average of 6 to 0.67 per worker per month. However, no specific e fect sizes were reported.

12. In this study, the effect of flextime was examined. The study design is a non-cont olled before–after study. This design is not very appropriate for measuring an effect, impact, or causal relation. The study has one serious weakness: the inclusion criteria were unclear.

13. a: Incorrect

    b: Reasonable

    c: Correct

    d: Reasonable

    The study design is not very appropriate for measuring an effect, and it contains one serious weakness, which makes the study prone to bias. We should therefore conclude that the trustworthiness of the findings is limited. In terms of p obability, we could say that the chance that flextime has a large positive e fect on tardiness is about 70%. This means there is a 30% chance that alternative explanations for the effect found are possible.

Did I get This 7.9

1. Yes: Correct

   No: Incorrect

   The abstract states that the researchers conducted "*a meta-analysis of the relationship of financial literacy and of financial education to financial behaviors in 168 papers covering 201 prior studies*." This means the study does not examine original, firsthand data collected by the esearchers themselves (primary research) but instead examines the results of previously published research (secondary research). This means that the study is a systematic review or meta-analysis.

2. Yes: Correct

   No: Partly correct

   In the Meta-analysis section, paragraph 2, the researchers state that they "*examined all studies that manipulated financial literacy with some education intervention or that measured financial literacy with well-known psychometric scales*." On page 5, they explain that "manipulated" refers to experimental and quasi-experimental studies, and "measured" to correlational and econometric ones. This means that the main study includes many kinds of study designs. In Figure 2 and section 2.2.2, however, it becomes clear that the researchers differentiate between results based on experiments (randomized controlled studies) and results of (three) other types of studies. This makes it possible to determine the effect of financial education on financial literacy based on only randomized con olled studies.

3. a: Correct

   b,c,d: Incorrect

   The appropriateness of a systematic review or meta-analysis that includes randomized controlled studies for measuring an effect, impact, or outcome is regarded as (very) high.

4. Yes: Incorrect

   No: Correct

   Section 2.1 Meta-Analysis Overview, paragraph 3, states that a computerized search in "numerous" databases was conducted using relevant keywords. It is unclear what "numerous" means (e.g., was a search conducted in ABI/INFORM, Business Source Premier, or PsycINFO?). In addition, the researchers did not contact experts to inquire about unpublished research, so it is not known if important, relevant studies were missed.

5. Yes: Incorrect

   No: Correct

   No information is provided as to how the selection process was conducted (e.g., whether the studies were selected by two independent reviewers), and there is no flowchart showing how many studie were excluded and why.

6. Yes: Correct

   No: Incorrect

   You can read in section 2.1 Meta-Analysis Overview (paragraphs 4, 5, and 6) which data were extracted and coded. A detailed overview of all extracted data and codes is presented in several tables in the appendix.

7. Yes: Incorrect

   No/unclear: Correct

   Assessment of the methodological quality of the studies included is not mentioned in the Meta-analysis section or in any of the other sections.

8. a,b,d: Incorrect

   c: Correct

   The Results section (subsection 2.2.1, first sentence) states that "financial education interventio have statistically significant but minuscule e fects: r^2 = .0011." According to the rules of thumb, an r^2 of 0.001 can be considered a very small (almost negligible) effect. In addition, Figure 2 shows that the effect sizes (r) are all in the range of .009 to .153, which can also be considered a small effect.

9. a: Correct

   b,c: Incorrect

   The Results section (subsection 2.2.1, first sentence), states that the e fect size of financial educatio interventions is r = .032 with a 95% confidence interval (C_{95}) of .029 to .035. The confidence interval in Figure 2 are also narrow, which means that the precision of the effects found is OK

10. This study examines the effect of financial education interventions on financial literacy. The stu design is a systematic review or meta-analysis based on randomizedcontrolled studies and other types of studies. This design is very appropriate for measuring an effect, impact, or causal relation. However, the study has three serious weaknesses: (1) It is unclear whether important, relevant studies were missed. (2) The process used to select studies was unclear. (3) The methodological quality of each study included was not assessed.

11. a: Not quite right

    b: Correct

    c: Not quite right

    d: Incorrect

    The study design is very appropriate for measuring an effect, but the study has three serious weaknesses, which makes it prone to bias. We should therefore conclude that the trustworthiness of the findings is moderate. In terms of p obability, we could say that the chance that financial educatio has a very small (almost negligible) effect on financial literacy is about 80%. This means that the e is a 20% chance that alternative explanations for the effect found are possible.

Did I get This 7.10

1. Yes: Incorrect

   No: Correct

When you read the abstract, you can see that data were collected through a case study of an employee-coaching program in a large, multinational construction company. This means that the study does not examine the results of previously published research (secondary research) but instead examines original, firsthand data collected by the esearchers themselves (primary research).

2. Yes: Incorrect

   No: Correct

   Despite the large amount of text that the researchers dedicate to defending their choice to use a single case-study approach, the last paragraph of the Research Design and Method section states that “the primary data were collected through a survey questionnaire.” In addition, you can see in Figures 1, 2, and 3 that the outcome of the questionnaire is expressed in frequencies and percentages.

3. Yes: Correct

   No/unclear: Incorrect

   In Figures 1, 2, and 3, you can see that only data collected after the attendance of the training program are provided. In addition, it is not stated in the Research Design and Method section or in any other section that the questionnaires were distributed and collected before and after the coaching took place, so we can safely conclude that all data were collected at the same time.

4. The study is a quantitative study in which all data were collected at the same time, which makes it a cross-sectional study.

5. a,b,c: Incorrect

   d: Correct

   A cross-sectional study is a design that lacks a control group, random assignment, and before and after measurements. Its appropriateness for measuring an effect, impact, or causal relation is therefore low.

6. Yes: Incorrect

   No/unclear: correct

   In the first paragraph of the Main Findings section, you can ead that the questionnaire was emailed to 83 employees. However, no information is provided about how these 83 employees were selected. This suggests that the selection of participants was not random, but rather a convenience sample of people to whom the researchers had easy access.

7. Yes: Incorrect

   No: Correct

   In the first paragraph of Main Findings, you can ead that the questionnaire was emailed to 83 employees, but the response rate was 67%, so the final sample size was 52. In the second-to-las paragraph of the previous section, however, it states that the coaching was provided at six annual events and that 40 to 50 employees attended each event. This means that in the 4 years preceding the study, more than 960 employees had followed the coaching program, so the sample size should have been larger than 52.

8. Yes: Incorrect

   No: Correct

   The last paragraph of the Research Design and Method section states that “*the survey was designed to investigate the impact of coaching on employees who had attended the programme. . . .*” As no references are provided for this questionnaire, we must assume that the researchers themselves developed it. Under The Attendees’ Perception of XYZ’s Coaching Program, you can read that users were asked in the questionnaire to rank the coaching program’s “*effectiveness in terms of enjoyment, the actual learning that occurred and its contribution to their personal growth*.” However, the researchers fail to explain what exactly “effectiveness in terms of enjoyment,” “actual learning,” and "personal growth" entail. In addition, no information is provided regarding the reliability and validity of these variables.

9. Yes: Correct

   No: Incorrect

   The researchers did not measure correlations or tested hypotheses but measured only the frequencies of the answers given by the participants.

10. a,b,c,d: Incorrect

    e: Correct

    As you can see in Figure 4, provides information about the number of participants who felt the coaching program had a low, slightly negative, neutral, moderate, or high effect on their organizational skills, but no effect sizes are given.

11. In this study, the researchers examine the effect of coaching on employee performance. This is a cross-sectional study, so its methodological appropriateness for measuring an effect, impact, or causal relation is very limited. In addition, the study has three serious weaknesses: (1) The sample was not randomly selected. (2) The sample was not large enough. (3) No reliable and valid measurement methods were used.

12. a,b,c: Incorrect

    d: Reasonable

    e: Correct

    The study design is not appropriate for measuring an effect, and the study has three serious weaknesses, which makes it very prone to bias. We should therefore conclude that the trustworthiness of the findings is (very) lo . In terms of probability, we could say that the chance that coaching enhances employees' personal and professional growth, motivation level, and performance is less than 60% or less. This means that there is 40% or more chance that alternative explanations for the effect found are possible.

CEBMa center for Evidence-Based Management

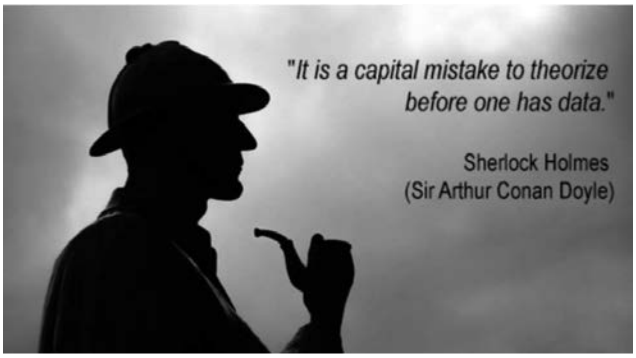
Open Learning Initiative

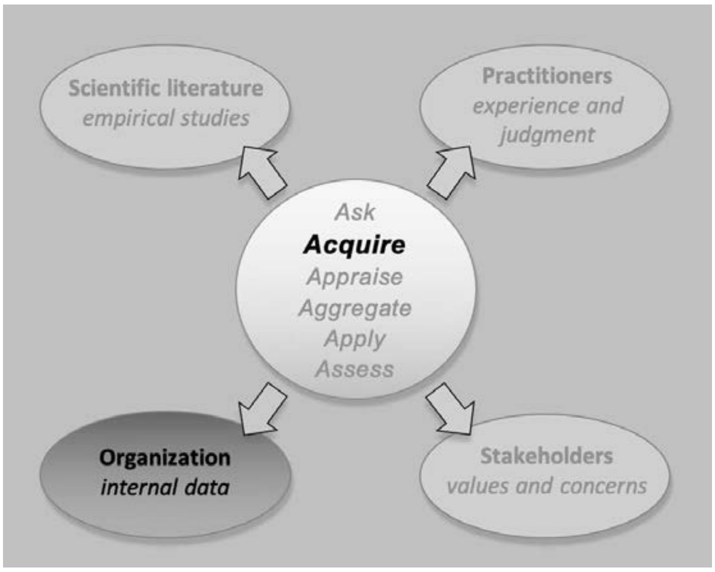
MODULE 8 | ACQUIRE Evidence from the Organization

Learning objective:

- Acquire evidence from the organization in a valid and reliable way.

Introduction





The modern organization continually generates data. From banks and hospitals to small retail stores, all organizations create data. Some data are created and recorded as part of regular operational activities, such as the appointment-booking system at a hospital or cash register at the local bakery. Some data help managers make decisions or monitor the organization's performance. Some data aid compliance with laws, regulations and accounting standards. Whatever their type or reason for creation, organizational data represent one of the richest sources of evidence for managers. Not only can organizational evidence be richer in volume or detail than evidence from the scientific literature, but it is also specific to the organizational context.

To use evidence from the organization, however, companies must have the skills to identify, combine and analyze data from multiple sources and possess the knowledge to build and apply analytical models to support decision-making. In this chapter, you will develop a better understanding of evidence from the organization and learn to acquire it in a valid and reliable way.

Why is evidence from the organization important?

As we have established, evidence-based management is about increasing the probability of making the right decision and understanding the degree of confidence you can place in that decision. In evidence-based management, managers and organization members make decisions within the context of a specific organization. Evidence from that organization is essential to:

1. Identify organizational problems or challenges
2. Determine organizational consequences of a problem
3. Recognize potential cause(s) of a problem
4. Find plausible alternative solutions
5. Monitor the effectiveness of management decisions or solutions

Learn by doing 8.1

Read the following scenario.

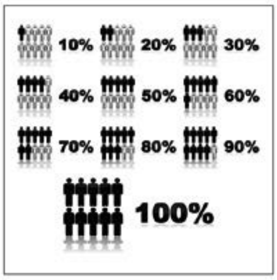
An Australian child welfare organisation notices that there is a serious shortage of foster parents. The CEO thinks this is due to the unfavorable economic situation, so she suggests increasing the financial compensation parents receive. An experienced childcare worker, however, thinks that foster parents drop out because there are more children with severe emotional and behavioral problems. He therefore suggests offering foster parents training that help them handle these challenges. Finally, a pediatrician thinks that the shortage can be attributed to the increase of children being placed into care.

How could evidence from the organization help? Check all that apply.

a) There may be data available about how serious and urgent the problem is (e.g. the size of the shortage)
b) There may be data available about the assumed increase of children with severe problems
c) There may be data available about the problem's organizational consequences.
d) There may be data available that suggest alternative causes for the shortage.

Systematically evaluating organizational evidence can be an effective and reliable way to verify claims regarding organizational problems. In addition, you can use data-analytic techniques such as regression analysis or predictive modelling to identify effective solutions. (1). In fact, consulting organizational evidence is often the best way to assess the impact of a management decision. After all, you make a management decision, and you can most easily judge its impact using evidence from your organization's own data. A general understanding of how evidence from the organization can be acquired and assessed is an essential skill for evidence-based managers (2).

From data to information to evidence

In common usage, the terms data, information, and evidence are used interchangeably. Data, however, refers to text, images, numbers or symbols that, if no context is provided, mean little or nothing to a human being. Bits of data are not meaningful without a context. Data can be regarded as raw material that can exist alone while information refers to data that have been processed in a meaningful and useful way. Only when you process bits of data (organize, structure, analyze and interpret), putting them in context, do they provide meaningful information. Taking these distinctions one step further, evidence only exists in the context of a claim or an assumption. Simply put, evidence is always evidence for (or against) something.

- **Data:** Numbers, words, figures, symbols, sounds, dates, images, etc. without context
- **Information:** Data processed in ways relating them to something or someone and considered meaningful or useful
- **Evidence:** Information supporting (or contradicting) a claim, assumption or hypothesis

Information technology and computer science emphasize the distinction between data and information. Computer systems are inherently full of electronic representations of words, numbers, symbols, etc. Unprocessed or out of any specific context, these forms of data a e meaningless.

For example, a list of numbers stored on a computer becomes “information” only when it is recognized as a list of the birthdates of employees. Put into context, the data becomes information when it answers such questions as “What is the average age of an employee?” or “How many people are likely to retire next year?” This means that if an IT professional refers to something as “data”, it is quite likely to require additional processing to make it meaningful or useful. In the context of this course, organizational evidence refers to data and the ways it has been transformed to make it more interpretable.

Did I get this 8.1

Of each item, determine whether it concerns data, information, or evidence.

1. A list of all employee identification numbers (EIN)
2. A list with identification numbers (EIN) of all employees older than 50 years of age
3. The total number of medical errors of an Italian hospital in 2020.
4. The total number of medical errors of two Italian hospitals in 2020, showing that one hospital had a 20 percent lower error rate than the other.
5. A graph based on the company’s annual financial performance showing a negative t end presented by the Chief Financial Officer who has repeatedly warned that the company’s revenues are decreasing.
6. The presentation of an Australian child welfare organization showing that more than half of the foster parents who stop foster parenting do so because they feel they received too little support.

What questions to ask?

In Module 2 it is explained that an evidence-based approach starts with asking questions. Setting out questions and framing diagnoses and problems correctly may be the most important steps in the evidence-based process. An evidence-based approach begins with the question “What is the problem, and what is the evidence for this problem?” or “What is the preferred solution, and what is the evidence that this solution will be effective?” Evidence from the organization can make the biggest difference in answering the latter part of these two questions. Acquiring evidence from the organization typically starts with a hypothesis or (assumed) causal mechanism. A hypothesis-led gathering and analysis of organizational data generates faster and more meaningful outcomes. It roots assumed problems (and preferred solutions) in “real” data relationships rather than correlations that were statistically significant by chance

Example

A large international insurance firm experiences a decline in profitability. Based on the assumption that its biggest clients are also its most profitable, the firm acquires organizational data regarding the number of services each major client uses and its profitability. The outcome, however, is counterintuitive: its largest clients turn out to be among the least profitable. Moreover, clients in the middle percentile, which do not require substantial resources, tend to use more services and are therefore more profitable than the larger clients at the top. The company therefore concludes that the initial core assumption is wrong. When this surprising outcome is discussed with its top executives, a new hypothesis emerges: the use of the company’s services is driven by client satisfaction and that usage determines a client’s profitability. To see whether this assumed causal mechanism is correct, the company acquires organizational data regarding client satisfaction, service usage and net profitability, and determines that a client’s satisfaction is indeed persistently and predictively linked to a client’s profitability. The company must now work out which variables drive satisfaction.

Some of the questions we identified as important in Module 2 have specific relevance to organizational evidence, including the following:

1. *What is the problem* (3)*: what, who, when, where?*
2. *Does organizational evidence confirm the problem?*
3. *Is there a trend: does the evidence suggest that the problem will worsen if nothing is done?*
4. *What organizational consequences of the problem does the evidence indicate?*
5. *Does the evidence confirm the assumed causal mechanism: is there a correlation between the assumed cause, the problem and its organizational consequences?*
6. *What is the preferred solution: what, who, when, where?*
7. *Does organizational evidence confirm the assumed causal mechanism: is there an association between the preferred solution and the favored organizational outcomes?*

Learn by doing 8.2

Consider this scenario that we used previously.

An Australian child welfare organization notices that there is a serious shortage of foster parents. The CEO thinks this is due to the unfavorable economic situation, so she suggests increasing the financial compensation parents receive. A childcare worker, however, thinks that foster parents drop out because there are more children with severe emotional and behavioral problems. He therefore suggests offering foster parents training that help them handle these challenges. Finally, a pediatrician thinks that the shortage can be attributed to the increased number of children being placed into care.

What evidence from the organization would you collect first?

a) Data regarding the number of children being placed into care and whether this number has increased.
b) Data regarding the number of children with severe problems and whether this number has increased.
c) Data regarding foster parents' satisfaction with the financial compensation they receive
d) Data regarding the size of the current shortage of foster parents and whether this shortage has increased.
e) The percentage of foster parents that have received training.

Learn by doing 8.3

Consider this scenario.

After acquiring organizational evidence regarding the assumed shortage of foster parents it is concluded that there is indeed a serious problem: the data show that the shortage is substantial and has rapidly increased in the past five years. As a result, both the number of children waiting for placement in a foster family and the average waiting time has dramatically increased. Thus, next step would be to identify the possible cause(s) for this problem. As mentioned, the CEO thinks the problem is caused by the limited financial compensation parents receive, whereas the childcare worker feels that foster parents drop out because they lack training to handle children with severe problems. Finally, the pediatrician suggests that the shortage may be caused by the increase of children being placed into care.

Based on this information, what evidence from the organization would you collect next?

a) Data regarding the number of children being placed into care and whether this number has increased.
b) Data regarding the number of children with severe problems and whether this number has increased.
c) Data regarding foster parents' satisfaction with the financial compensation they receive
d) The percentage of foster parents that have received training.
e) Nothing, I first need more information.

Note that organizational data may also be gathered in anticipation of future needs, as opposed to current problems. For example, as part of a project to build evidence-based management capabilities, an organization might build a data warehouse. Data warehouses are repositories that integrate data from multiple sources across the organization. Such data might be gathered as a basis for future hypothesis testing and learning as the organization's capacity to ask critical questions increases.

What types of evidence are typically available?

Evidence from the organization comes in many forms. It can be derived from financial data such as cash flow or costs, from customers in the form of customer satisfaction or from employees through information about retention rates. Evidence from the organization can be “hard” numbers, for example staff turnover rates, medical errors or productivity levels, but it can also include “soft” elements such as perceptions of the organization’s culture or attitudes towards senior management. There are many ways to classify organizational data and the evidence it produces, including these broad types.

Finance and accounting

Historically, managers have paid considerable attention to financial and accounting data. Many events in the organization, such as the sale of a product or the delivery of a service, generate data relevant to the organization’s financial position. The sale of a product will be represented by entries in one or more sets of records called ledgers. Organizations use data from ledgers to create key financial information, such as

- Statements of cash flow – a record of money received or given out
- Income statements – lists of an organization’s profit or loss and income
- Statements of a firm s financial position (also known as a balance sheet) – lists of an organization s assets (money or things they own) and liabilities (money or things they owe)

Financial data use typically requires some understanding of financial and accounting concepts and thus may require specialized professional help with interpretation.

Human resources

Human resource evidence is fundamentally about people: who they are, their characteristics and their relationship to the organization. Examples are pay, grade, tenure, years of experience, attendance, job satisfaction, engagement and performance. They may also include staff surveys, policies and data regarding the activities an employee carries out.

Sales and marketing

Sales and marketing evidence includes facts about the number of products or services sold, market share, competitors, details of customer relationships, brand awareness and the impact of marketing campaigns.

Risk

Large companies typically have departments that manage and assess the multiple risks that can impact the organization. Perhaps the most developed risk functions are in banks and insurance companies, where organizational risk is managed and evaluated in terms of operational risk, market risk and credit risk. Evidence about risk may take the form of calculations about the potential risks to an organization based on its current state or strategy, but it may also entail a detailed analysis of what has gone wrong during standard operational procedures.

Production

Production evidence relates to the products or services that an organization creates, including measures of inputs, outputs and the overall quality level. Those outputs may be physical objects such as cars, personal services such as haircuts or intangible services like legal advice. Note that some organizations may use the term “operations” instead of “production”.

Quality

Larger organizations often capture data to monitor, control and ensure the quality of their products or services. Especially high-reliability organizations in the healthcare, aviation, petrochemical, food or banking industries have complex quality-management systems that closely monitor data such as accuracy, timeliness, failure frequency rate, safety, and other Key Performance Indicators (KPIs, see later in this chapter).

Customer service

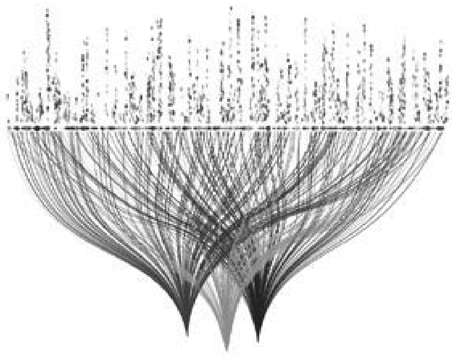
Specific customer service functions deal with client interactions that do not involve selling or production. For example, they produce data regarding the number and content of customer complaints.

Learn by doing 8.4

Determine the place where you most likely could obtain the following types of evidence:

1) The average age of the employees of a small-to-medium sized company.
2) The average number of monthly complaints a large telecommunications company has received in the past year.
3) The number of patients admitted to the Intensive Care Unit (ICU) of a large academic hospital in a specific month.
4) The current number of patients admitted to the Intensive Care Unit (ICU) at a large academic hospital.

Big data



“Big data”, a particular kind of organizational data, is a hot topic in contemporary business. The concept, however, is still nascent, and, as a result, many definitions exist. Most of these definitions have the following “three Vs” in common:

1. Volume

Size is what first comes to mind when we think of big data. As a rule, big data is comprised of multiple terabytes or even petabytes of structured and unstructured data. To give you some idea of what this volume is, one terabyte stores as much data as 1,500 CDs or 220 DVDs, the equivalent of about 16 million Facebook photographs (4). One petabyte equals 1,024 terabytes, which is enough to store the DNA of the entire population of Europe.

2. Variety

Big data come from a variety of sources. Organizational data are acquired from sources such as management-information systems, personnel systems and physical records (paper spreadsheets, for example). This is referred to as “structured” data. But nowadays data also come from emails, text messages, tweets, audio recordings, photographs, videos, etc. – this kind of data is unstructured data. Technological advances allow organizations to store, process and analyze these types of data, and thus use it for economic purposes

3. Velocity

Velocity refers to the speed at which data are generated, leading to a growing need for real-time analytics. Nowadays even “conventional” retailers generate data at a tremendously high rate. Walmart, for instance, processes more than one million transactions per hour (5). Data also are pouring into organizations with increased speed due to “firehose” data sources such as social media (6).

These three characteristics of big data raise questions such as: How can I accurately analyse 100+ terabytes of heterogeneous data per day in real-time? For this reason, new data analytics using artificial intelligence, machine learning, and neural networks have emerged, suited to big data’s complexity.

> ### Did I get this 8.2
> **Read the two scenarios and determine whether it concerns big data.**
>
> ***Scenario I***
>
> *The integrated database of a large American life insurance company with assets amounting to approximately 578 billion U.S. dollars contains the records of more than 45 million customers.*

The database not only contains customers' contact information (name, address, phone number, email address, etc.), demographic information (age, gender, marital status, income, education, employment, etc.) and information regarding their life insurance policy (coverage, term, conditions, costs, beneficiaries, etc.), but also information such as medical history, risk profile, the number of contacts with the call center, customer profitability, etc. On average the database contains about 120 datapoints per unique customer record.

Scenario II

A global market research data and analytics company is specialized in the measurement of what global consumers watch (online, TV, radio, books, magazines, newspapers) and buy (grocery, drugstore, convenience stores, mass retailers). As such they collect 1.7 billion TV viewing records and 6.7 billion store transactions each month. By combining this information with millions of other consumer attributes (such as demographic, geographic, behavioral, and personality variables) they can provide insight into consumers' current (and future) media and buying behavior.

a) Both scenarios can be considered examples of big data
b) Only scenario I can be considered an example of big data
c) Only scenario II can be considered an example of big data.
d) None of the scenarios can be considered an example of big data.

Where to find organizational evidence

Organizational evidence and the data that generate it can be found both inside and outside the organization. The types of organizational evidence that we discussed earlier and the examples we provided in the Learn By Doing exercise can be a good starting point. For example, the finance department is the key custodian of the organizational and departmental budgets and is therefore often the most important source for acquiring financial data. However, data may also be stored at locations outside the department that uses it. For example, human resources (HR) data are found in the HR department but also in the finance division and in local branches of the organization. It is therefore important to gain an understanding of how and where your organization captures and stores its data. Below is an overview of the most common places to find organizational data.

Databases and information systems

Databases usually are the core systems for capturing and processing many of the organization's daily activities. In general, a database consists of data structured into records of individual elements. For example, an address record may contain the house number, street name, city and postal code. Most databases link together different types of records using a common identifier – a case in point is employee salary records and employee addresses. These data may be kept in different records linked by a common identifier (such as staff ID or personnel number) stored in both. Identifiers also link data f om different databases. Most information systems within an organization – such as management-information systems, personnel systems or customer-relationship management systems – consist of an underlying database you can query directly, with tools for searching, extracting and analyzing its data. Sometimes the data in these separate information systems is uploaded into one large, integrated database, also referred to as a data warehouse.

Document- and content-management systems

A great deal of organizational data is stored in the form of documents or spreadsheets rather than as structured data in a database or data warehouse. There is a wide variety of systems for managing documents including document-management systems and systems for managing intranet/Internet content. These types of systems provide several functionalities for classifying and searching for documents.

Workflow systems

These are systems that manage the execution of a business process. Workflow systems a e often hybrids between data and document-based systems: generating and storing both data and whole documents.

Physical records

Many organizations still use physical records, including documents in filing cabinets. This could be because of lack of funds to invest in new technology, legal requirements or simply habit. Physical records can present a challenge if they are a relevant source of organizational data. There are a variety of techniques for turning physical records into more easily analyzed organizational data, from scanning/optical character recognition to physically reviewing and screening documents. Depending on the balance between the potential value of physical records and the additional costs of analyzing, they may still be of great value. In fact, in some cases the explosion of digital data means that it is sometimes easier and quicker to find the relevant organizational evidence in physical rather than electronic format.

Staff

A great deal of organizational evidence exists at staff level. For example, relevant data may be kept by individual staff members on their own PCs, on shared drives or in the form of physical records (e.g. physicians keeping patient records). As a result, senior management may be unaware how subordinates actually store and maintain data.

Industry bodies, professional associations and census bureaus

Industry bodies, professional associations and census bureaus often have relevant, high-quality information about an organization, its competitors and industry or sector. Often such organizations provide valid and reliable information about how a company's metrics and KPIs compare to the average in the sector.

Social media

Relatively novel, but increasingly important, sources of organizational evidence are social media sites such as Facebook and Twitter. Not only do social media generate data regarding customer satisfaction, brand awareness, brand identity or perceived quality, but they also provide information about the organization's relationship with society.

External stakeholders

External stakeholders such as customers, regulators, shareholders, suppliers and even the society at large may be an important source of evidence about a specific organization. These external stakeholders are a rich source of opinions about the output, results, interactions and behavior of an organization.

Did I get this 8.3

Read the following scenario.

The dean of a Dutch community college for vocational education (a school that that prepares 16 to 18 year olds to become an electrician, hair dresser, mechanic, etc.) noticed that during information sessions, parents often ask questions about student safety (e.g. bullying, harassment, intimidation, abuse, verbal and physical aggression, alcohol, drugs, etc.). She usually states that the number of safety related incidents is very low, but this answer is based only on her personal experience rather than organizational evidence. Given the importance of the topic, she therefore decides to consult the school's management information system. Although teachers and students are required to report all safety related incidents, the school's management information system indicates that in the past semester there were only 12 safety related incidents. Because she considers this number to be quite low, she decides to consult additional sources of organizational evidence to see whether this low number is likely to be true.

What would be the source that MOST likely would have information on the number of safety incidents?

a) Physical records

b) Staff

c) Industry bodies, professional associations and census bureaus

d) Social media

Techniques and tools for acquiring data

There is a range of techniques and (software) tools available for acquiring the organizational data that can become evidence for managerial decisions. Below is a brief overview of the most common.

Management information and business intelligence systems

There are many specialized tools available to extract data from both core processing (operations) systems and specifically designed reporting systems. These are commonly referred to as management information or business intelligence systems. The purpose of these systems is to support the decision-making process by organizing, processing and analyzing the data and turning them into useful information. Such systems often comprise multiple components, such as an ETL function (extract, transform, load – the process of extracting data from source systems and bringing it into the company's data warehouse), and a separate database along with reporting and analysis software. In addition, they often include advanced dashboards and visualization tools that present data in graphs, charts and accessible visualizations.

Querying existing databases and systems using SQL

Most information systems store data in an underlying database (i.e. data warehouse). It is possible to "query" the database, extracting data from the database in a readable format according to the user's request, by using what is known as Structured Query Language (SQL). Writing basic queries in SQL, however, is a technical skill that requires the support of the organization's IT department.

Analyzing organizational data using statistical software

Choosing statistical software and/or a data-analytics tool is a trade-off between costs, benefits and ease of usage. Popular packages are SPSS, SAS, R, Python and Excel. SPSS is a user-friendly package offering a range of data-analytical functions. The downside is it is costly. SAS is a good choice for a person who analyses complex data sets daily. However, if you use statistics program only once or twice a month, it may not be worth spending hundreds of hours learning SAS language. The same counts for R and Python, two popular programs in the data-science world. Both have a command-line interface, requiring code or scripts, making them labor-intensive to master. R is open source and downloadable for free. Finally, Microsoft's Excel too provides a wide range of statistical functions. Excel is easy to learn and straightforward to use, and it produces attractive graphs and story-telling charts.

Review of documents and reports

A document review can be a quick and easy method of uncovering useful organizational data. It can be done with or without the assistance of specialist software such as a program for content analysis or text mining.

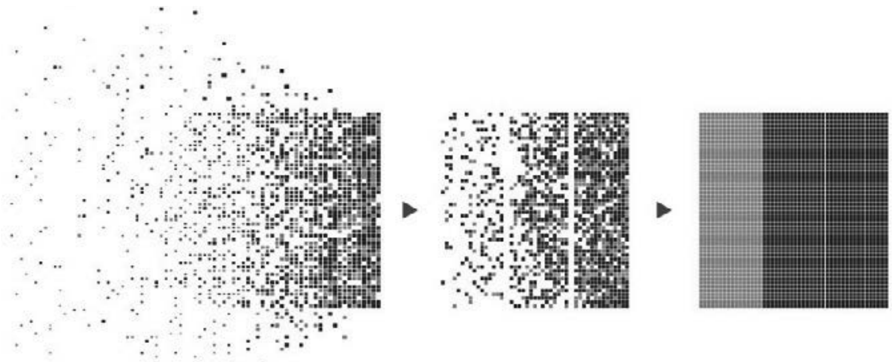
Surveys

A common form of acquiring organizational data is the survey. As explained in Module 3, effective survey design requires training to avoid bias. Probably the most popular survey program is SurveyMonkey, a free online tool you can use to construct a questionnaire (based on a customizable template), collect responses and analyze results.

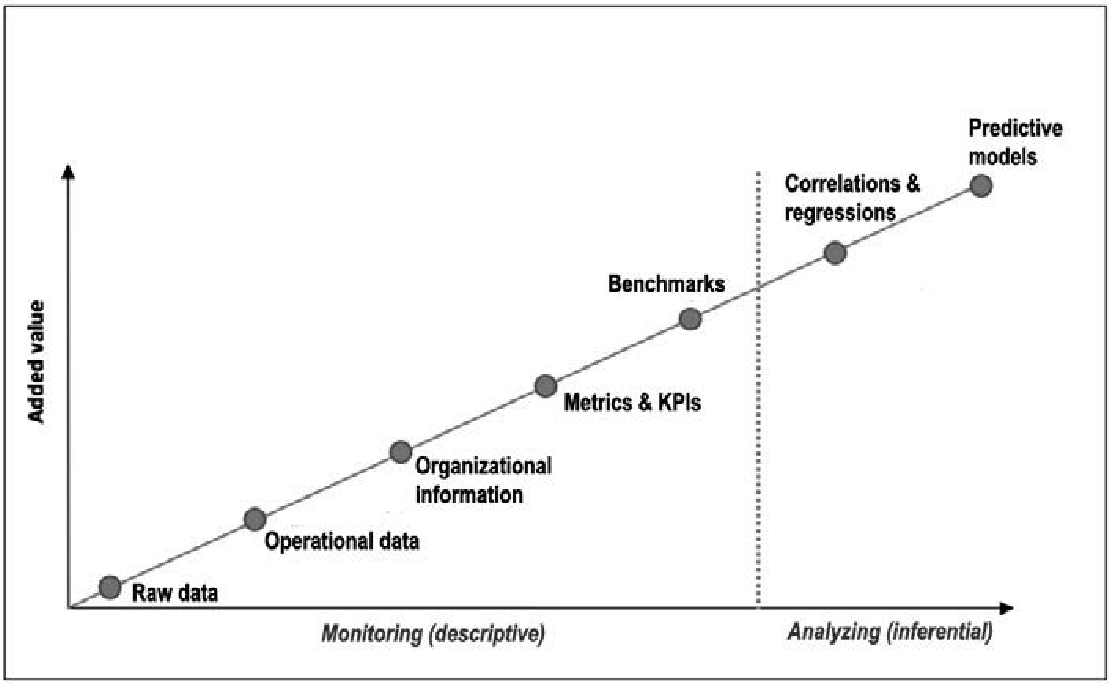
Example

In the last 'Did I Get This' exercise we used the scenario of a Dutch community college for vocational education. The school's management information system indicated that in the past semester there were only 12 safety-related incidents. Because the dean considered this number to be quite low, she decided to acquire additional organizational evidence. This scenario is based on a real life example. The dean decided to run a survey among the schools' teachers and a random sample of students through SurveyMonkey. The outcome confirmed that the number of safety-related incidents at the school indeed was indeed rather low. This outcome was also confirmed by Dutch census data, indicating that in the region where the school is located, most community colleges experience few safety incidents.

Organizational data turned into information



Besides categorizing organizational evidence based on the function or physical location of its original data, organizational evidence can be thought about in terms of the value it adds to the organization's decision making. In the Figure below (7) , the vertical axis represents the level of added value, a composite index of the degree of interpretability and uncertainty reduction that various forms of organizational evidence provide. Typically, raw data add the least value because without some degree of processing they are difficult to interpret. In contrast, predictive models can indicate the importance of current evidence for future outcomes, for example the effect of employee turnover rates on important organizational outcomes like safety or service quality. The horizontal axis indicates the nature of the analyses performed on the organizational data. Descriptive evidence is used to monitor organizational evidence relative to past facts (last year's sales) or current environmental conditions or standards (benchmarks). Inferential evidence transforms descriptive evidence through analytic techniques into indicators of association (e.g., the correlation of employee demographics with turnover) or predictive models (e.g. regression modelling used to forecast future turnover from present employee demographics and other evidence). As you can see, we expect greater added value from inferential analyses of organizational evidence than from descriptive evidence used for monitoring.



Descriptive organizational evidence

Operational data

Organizations create large amounts of operational data during their everyday activities – even if it is not always easy to extract, aggregate, or interpret those data. Operational data are collected automatically within systems related to production, sales, services delivered, personnel, customer service and many other routine functions. Operational data are descriptive in nature: They describe what is going on in an organization. Sometimes operational data are “real time”, providing a snapshot of the present situation, but more often they are retrospective, providing a picture of how things were in the past (like a rear-view mirror). Collecting operational data enables an organization to identify general trends, typically expressed as frequencies, averages (means, medians, modes), proportions or ratios.

Organizational information

Although operational data can provide managers with useful information, the added value is often low. For example, an individual sales transaction may not give a manager much help in making decisions. However, combining all the sales for a month into a sales report can make it possible to make judgments and identify trends. Aggregated operational data presented within a functional context (e.g. sales, finance or HR) a e referred to as “organizational information”. Within most companies some form of organizational information is available about:

1. Staff
2. Physical assets e.g. building, equipment, inventory
3. Clients / customers / patients
4. Financial assets (e.g. bank balances) and liabilities (e.g. debts)
5. Business/operational processes
6. Marketing and sales
7. Production / operations / performance
8. Quality / safety

Although you can find certain organizational information in readily identifiable systems, much of it is likely to be dispersed across the organization, especially in large and complex firms. Organizational information is, like operational data, descriptive in nature. It is, however, beneficial for developing a general understanding of how the company currently performs, and identifying changes in the organization’s performance over time.

Metrics and Key Performance Indicators (KPIs)

A metric is a measure. An organization’s performance metrics can be a single measure or derived from a combination of two or more measures. Examples are financial ratios (e.g. leverage, profitability, total revenue over time), daily occupancy rate, number of medical errors per 1,000 patients or average handling time per phone call. When we tie a metric to a target, goal or norm critical to the organization’s success, we often refer to it as Key Performance Indicator (KPI). Although many metrics are presented as numbers, KPIs may take the form of ratings on a scale such as the RAG status (for red, amber, green), with red meaning that there is a severe problem and corrective action is required, amber indicating a potential problem, with the situation needing to be monitored closely, and green signifying that the performance is on target and no action is needed.

KPIs tend to add more value than operational data and organizational information because they are applied relative to targets or guidelines, that is, their use is contextualized. Typically, KPIs are measured in the context of performance goals for individuals or units, and provide information regarding both past performance and goal progress. KPIs can also function as performance guidelines specifying minimums and maximums. They can operate in settings with minimum guidelines, as in the case of the number of patients a physician needs to see a week in order to cover practice expenses. Or they may be maximums such as capacity limits as in the case of the number of clients an account manager can effectively serve. KPIs are used to determine whether the organization's performance has changed over time, and, if so, at what rate and in which direction. In addition, KPIs are a convenient shortcut to identify problems, or to determine the impact of management decisions.

However, as we will discuss in the next module, the existence of a colorfully presented and precise KPI does not necessarily mean that the underlying measures are relevant or based on good-quality data. In fact, not all KPIs are based on hard numbers, but may simply be a subjective judgment.

Benchmarks

These are metrics that are tied to standards or best practices within the industry. Thus, "benchmarking" is a method of systematically improving a company's performance by measuring and comparing its performance against an organization that is considered to be the "best in class". For example, in healthcare, there are industry data indicating "door-to-needle time" for treating stroke victims. A hospital manager would want to acquire information on door-to-needle time in the hospital's emergency department in order to learn whether it should be improved. In general, benchmarking does not just involve the collection and comparison of benchmarks, but also further data-analytics to identify the underlying cause of underperformance and actions for improvement. Benchmarks can be rich sources of information that can help managers determine how the organization measures up to others in the sector. In fields such as healthcare, benchmarking has been encouraged by setting national standards and by the publication of hospital performance metrics.

A caveat regarding benchmarks is warranted. You need to critically evaluate whether the benchmarked organization is truly "best" on some criterion. Don't confuse hype or publication relations with actual evidence of success. Copying what others do only makes sense if you know that what was done was effective and is likely to be appropriate in your context. The expression 'best practices' implies that these practices are best for most organizations and that there is good evidence to support this claim: Both claims are unlikely.

Did I get this 8.4

What type of organizational evidence are the following descriptive measures?

1. The number of employees leaving the organization *in the past 6 months*
   a) Operational data
   b) Organizational information
   c) Metric
   d Key Performance Indicator
   e) Benchmark
2. A company's number of sales
   a) Operational data
   b) Organizational information
   c) Metric
   d) Key Performance Indicator
   e) Benchmark
3. The number 20, representing the percentage below target of vacuum cleaners sold by a sales agent in the past month.
   a) Operational data
   b) Organizational information
   c) Metric
   d) Key Performance Indicator
   e) Benchmark

Inferential organizational evidence

Correlations and regressions

Two metrics are correlated when a change in the value of one metric leads to a change in the value of the other – a matter we have discuss in detail in Module 6. For example, temperature and ice cream sales are correlated: when the temperature increases, ice creams sales go up as well. Regression concerns the prediction of an outcome metric from one predictor metric (simple regression) or several predictor metrics (multiple regression). For example, for every one-degree rise in temperature, about 1,200 more ice creams are sold on average.

Correlations and regressions are inferential measures. Whereas descriptive measures such as operational data and KPIs provide rather basic information about the organization's performance, inferential measures are produced by statistical calculations that allow us to infer trends, identify cause-and-effect relationships and make predictions about organizational outcomes. However, as we will see in the next module, inferential measures can be misleading, and therefore require the highest degree of scrutiny.

Predictive models

A predictive model is a statistical model we use to make a prediction. The term, however, is subject to debate. Some use it in a generic sense to refer to any statistical model that is used to make predictions (such as correlations and regressions), whereas in the realm of "big data" it often refers to a model that uses complex algorithms derived from advanced data-analytic techniques based on artificial intelligence or machine learning. Predictive models are powerful tools in the decision-making process: they can identify drivers and predictors for outcomes relevant to the organization. Although the possibilities of predictive models such as artificial intelligence and neural networks are exciting, all the usual principles of evidence-based management apply: They are just another tool in your toolkit, and their use requires a critical mindset and attention to the quality of the evidence used to construct and populate the model.

Did I get this 8.5

At the beginning of this page, we have discussed the terms "descriptive" and "inferential" evidence. What is the difference between these two?

a) In general, inferential evidence adds more value than descriptive evidence.

b) Inferential evidence uses analytic techniques to transform descriptive evidence into associations and/or predictive models

c) Descriptive evidence is used to monitor past and/or current phenomena (e.g. sales, absenteeism) whereas inferential evidence can be used to forecast future phenomena

d) All of the above

Articles on data analytics often refer to analytics in terms of machine learning, neural networking and artificial intelligence. These tools are useful but the real value in data analytics comes from simple statistical techniques that can be deployed by most managers. If you have access to Excel, you can work out means, trends, correlations and regressions, and if you don't know how to, there are some helpful short tutorials on YouTube. The analytical tools are important, but even more so is the mindset of evidence-based management – the attitude of "Hey, we don't know the answer. Let's get some data and find out!"

Some final considerations when acquiring data

The hype of the current era regarding data analytics and Big Data could give the impression that organizational data are always readily and cheaply available. This, however, is often not the case. In fact, many organizations struggle to manage and fully utilize their data. In addition, many data-analytics projects in organizations have started with high expectations but failed to deliver on their promises, because fundamental issues in data usage were not considered. Some of these issues might include:

Data protection/information security

There are many laws and regulations aimed at keeping employees' sensitive information safe and the privacy of customers secure. These include the Data Protection Act (2018) in the United Kingdom, the General Data Protection Regulation (2016) in the European Union, and the Federal Information Security Management Act (2002) in the United States. We need to consider the impact of these laws when using organizational data. Multi-national firms in particular may be impacted by different, sometimes conflicting, regulations. In some cases, this makes it difficult to use data from multiple locations. In other cases, we need to obscure some of the data (e.g. replacing names and addresses of customers with codes). Working with organizational data means negotiating the data-protection rules. This means that as managers we need to think carefully about the type of organizational data we require (need to have) and what data we can exclude (nice to have) without reducing the value of the evidence we have in hand.

Costs vs benefits

Acquiring and analyzing organizational data may come at a considerable price, so a thorough assessment of the expected costs and benefits is important. Several large companies have invested millions of dollars in building data warehouses or implementing big data projects where the quality of the data fed into the systems was so weak the cost expended was unwarranted.

Accessibility

While some organizational data may be captured and stored in an easily accessible central database, much of it is likely to be dispersed across the organization, especially in large and complex companies. Acquiring data from multiple systems and locations may require a lot of time and effort. In addition, when we lack a common identifier, it will be very hard to link data from different systems. For this reason, we often require assistance from an internal IT department or an external data-analytics expert.

Organisational politics

Finally, organizational politics sometimes stand in the way of successful use of organizational evidence. Even when we capture and store data in an accessible database and have sufficient knowledge of data-analytics we still may face challenges in using organizational evidence to support decision-making. In fact, lack of executive sponsorship is consistently cited as the main reason data-analytics projects fail. One of the reasons is that evidence-based management sometimes challenges authority. It brings in evidence that may contradict the intuition and judgment of business leaders. For example, when the board has decided to move the company's customer-service operation overseas, and you want to test the hypothesis that profitability has dropped due to a decrease in customer satisfaction, you should not be surprised if some managers object to gathering relevant evidence. As a data-analyst once stated: "At day's end, we depend on data and technology, but we live and die by politics." (8)

Learn by doing 8.5

Consider the following scenario.

Senior executives in a global investment bank with more than 6000 locations are concerned with the number of transactional errors made in the back office – in the past year this involved a total amount of 120 million US dollars To get a better understanding of the nature of the problem, they want to collect organizational evidence, such as the number of daily errors made in the past six months, whether there is a trend, in which locations most errors were made, whether there is a relationship with the locations' absentee rate, the average education level and tenure of the employees involved, and the performance rating and average years of experience of their supervisors

What could be a potential barrier to acquiring this organizational evidence?

a) Costs vs benefit
b) Accessibility
c) Data protection and privacy regulations
d) Organizational politics

We started this module by explaining why evidence from the organization is important: It helps you to verify claims regarding organizational problems, determine its organizational consequences, recognize potential cause(s), and possible solutions, and assess the effectiveness of these solutions.

We then discussed the terms “data,” “information,” and “evidence.” In daily life these terms are used interchangeably, but in the realm of evidence-based practice and information technology there is a clear distinction. Data refers to text, numbers or symbols that have little meaning without a context. When the context is provided data become (meaningful) information. When information is related to a claim or an assumption it becomes evidence. Simply put, evidence is always evidence for (or against) something.

Next, we discussed which questions to ask. Relevant questions are ‘What is the problem (what, who, when, where)?’, ‘Does organizational evidence confirm the problem? ’ ‘Is there a trend?’, ‘What are the organizational consequences of the problem’, ‘What is the preferred solution’, and ‘Can organizational evidence be used to monitor the effectiveness of the solution?’

There are many ways to classify organizational data and the evidence it produces, such as financial-, HR-, sales and marketing-, risk-, operational-, customer-, quality-, and performance data. This list can also be a good starting point to finding relevant data.

However, HR data are typically found in the HR department but also in the Finance division and in local branches of the organization. It is therefore important to gain an understanding of how and where your organization captures and stores its data. Large organizations often have advanced integrated information systems, where small and medium-sized organizations often store their data in separate databases such as personnel systems and customer-relationship management systems. Important sources that are often overlooked are staff, external stakeholders, professional associations, and census bureaus.

Next, we discussed ‘Big data’, a particular kind of data that have the “three Vs” in common: large Volume (multiple terabytes), from a Variety of sources, and Velocity – the speed at which these data are generated and change.

At the end of this module, we discussed how organizational evidence can be thought about in terms of the value it adds. Descriptive evidence is used to monitor current organizational evidence relative to past facts. Examples – in order of added value – are operational data, organizational information, metrics, KPI’s and benchmarks. Inferential evidence on the other hand transforms descriptive evidence into indicators like correlations or predictive models. In general, inferential evidence tends to add more value than descriptive evidence.

Finally, when acquiring organizational data, we should take into consideration aspects such as privacy, data-protection laws, the cost-benefit ratio of acquiring and analyzing organizational data, and internal politics that sometimes stand in the way of successful use of organizational evidence.

Podcast: ACQUIRE Evidence from the Organization

In this podcast host Karen Plum discusses with Martin Walker, Director of Banking & Finance at CEBMa, Jeroen Stouten, Professor of Organisational Psychology at KU Leuven University, and Eric Barends, Managing Director of CEBMa, evidence gathered from the organisation to aid our multi-source evidenced-based decision making. The appraisal of this evidence is covered in Module 9 and its corresponding podcast episode.

In this podcast we consider some of the things we need to think about before either using existing organisational data or collecting data specifically gea ed to address our question or area of interest. As with other sources of evidence, there are a myriad of obstacles and roadblocks to be aware of in our search for organisational data and information to inform our decision making.

There is a tendency for people to place organisational evidence in high regard, but as we will see, this faith can often be misplaced – not for nefarious reasons, but in some cases simply due to a lack of understanding when designing systems, or using data that has been churned out for years without anyone just asking where the data comes from.

As ever, we keep asking questions!

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

References

1. Note that organizational data is also a valuable source of evidence for scientific inquiry in management Researchers need to be able to work with organizational data too.

2. Barton, D., & Court, D. (2012). “Making advanced analytics work for you.” *Harvard Business Review*. Volume 90. Number 10. 78-83 Pages.

3. Of course, evidence-based management is not only about assumed problems but also about opportunities. In case of a (perceived/assumed) opportunity the questions above can be rephrased accordingly.

4. Gandomi, Amir, and Murtaza Haider (2015) Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35 (2)

5. Cukier K, 2010 Data, data everywhere: A special report on managing information, *The Economist*, 25 February. Retrieved from http://www.economist.com/node/15557443. .

6. Dykes, B (2017) Big Data: Forget volume and variety, Focus on velocity. Forbes, 28 June. Retrieved from https://www.forbes.com/sites/brentdykes/2017/06/28/big-data-forget-volume-and-variety-focus-on-velocity/.

7. Cascio, W, & Boudreau, J (2010) *Investing in People: Financial impact of human resource initiatives*, Ft Press.

8. Vaughan, R (2015) *Politics over analytics: BI and Data Science findings need a pitch*. LinkedIn, 28 January. Retrieved from https://www.linkedin.com/pulse/politics-over-analytics-bi-data-science-findings-need vaughan-robison/.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 8.1

a,b,c,d: Correct > In fact, all four options are good examples of how the organization's data could help. Often, organisational data can help determining whether an (assumed) problem is serious, urgent, and has negative organizational consequences - and thus indeed requires managers' attention. In addition, organisational data may confirm (or efute) the assumed cause(s) for the shortage or suggest alternative causes. For example, in this case there may be data available based on exit-interviews about the reason why people no longer want to be a foster parent.

Did I get this? 8.1

1: Data > If no context is provided, a list with employee identification numbers have little meaning and thus should be considered data.

2: Information > In this case the data (EIN) are presented in a meaningful context – representing all employees older than 50 - and thus should be considered information.

3: Information > The numbers are provided in a meaningful context - medical errors in a hospital – and should therefore be considered as information.

4: Information > The numbers are provided in a meaningful context but do not support (or contradict) a claim, assumption or hypothesis, and should therefore be considered information.

5: Evidence > The data (the graph) is presented in a meaningful context - the company's financial performance – and are tied to a claim – the company's revenues are decreasing – and should therefore be considered as evidence.

6: Information > Although the data are presented in a meaningful context and even provide a plausible explanation for why foster parents drop out, they are not tied to a claim, assumption or hypothesis, and thus should be considered information.

Learn by doing 8.2

a,b,c,e: Incorrect

d: Correct

When taking an evidence based approach the first questions to ask a e Does the organizational evidence confirm the assumed p oblem? and Is there a trend – does the evidence suggest that the problem will worsen if nothing is done? After all, if the organizational evidence does not confirm the e is a serious shortage of foster parents, there is no problem to solve, and further collection of evidence is pointless.

Learn by doing 8.3

a,b,c,d: Incorrect

e: Correct

This is a bit of a trick question. Although all causes may sound plausible, they are suggested by only one person. In addition, it is unclear how trustworthy their judgments are. As discussed in Module 4 (Appraise: Evidence from Practitioners), even the judgement of experienced professionals is prone to cognitive bias. Thus, more information is needed. For example, do other practitioners agree with the suggested cause(s) of the problem or do they see plausible alternative causes? As explained, acquiring evidence from the organization starts with a hypothesis. Ideally, this hypothesis is supported by evidence from other sources, as this prevents a 'fishing expedition' in which ir elevant data is captured and pointlessly analysed.

Learn by doing 8.4

1: Information regarding the average age of an employee can typically be found at the HR department. Sometimes, however, this information can also be found at the Finance department. Large companies often have a special 'people analytics' team or department that can provide you this information and run additional analyses.

2: Although it is not unlikely that the marketing managers and sales agents are aware of the number of complaints, information regarding the average number of monthly complaints can typically be found at the department in charge of customer services. However, some large companies have an integrated information system – also referred to as data warehouse – that can provide you with all the information that is collected within the organization.

3: Information regarding the number of patients admitted to the ICU in a particular month is most likely available at the patient administration department – the department that is responsible for all patient accounting, registration, admissions, discharges, medical records, and other patient-generated information. However, this information may also available at the Finance department.

4: This is a bit of a trick question. The Patient Administration department is responsible for all patient accounting, registration, admissions, discharges, medical records, and other patient-generated information, so this would be the place to obtain information regarding the number of patients admitted to the ICU. However, it takes some time before data/information are collected, processed and made available, so information regarding the *current* number of patients admitted to the ICU would typically be available at the ICU department itself.

Did I get this? 8.2

a: Partially correct

b: Incorrect

c: Correct

d: Partially correct

In the second scenario billions of records are combined with millions of other types of information results in multiple terabytes of unstructured data. In addition, the data are 'dynamic': they change on a daily basis and each month a large amount of new data are generated. So you are correct that this can be considered big data.

However, the first scenario can not be conside ed big data. Even with 120 datapoints per customer record the total number of data would equal 120 x 45 = 5,5 billion data, which would easily fit on a single DVD. In addition, these data are relatively 'static': they won't change much over time, and only a limited amount of new data are generated.

Did I Get This 8.3

a: Incorrect > It is not likely that there are physical records regarding the number of safety incidents.

b: Reasonable. You are correct that staff, in particular the teachers, would know about safety related incidents. However, this information is not (yet) available, so an option would be to survey students about their experiences – we will discuss this option in the next page.

c: Correct > It is likely that census bureaus, unions, industry bodies and/or professional associations hold data regarding the safety and number of incidents at community colleges, often organized by region to make comparison or benchmarking possible.

d: Incorrect > Although some students (or parents) may turn to social media when they feel that their concerns are not being heard, it is not likely that this will provide reliable evidence regarding the schools student safety.

Did I get this? 8.4

1c: Correct.

1a,b,d,e: Incorrect

The data (number of employees) are presented in a meaningful context (who have left the organization) and are combined with another measure (the past six months). However, there is no connection with a performance goal or best practice, which makes it a metric.

2a: Correct

2b,c,d,e: Incorrect

The data (number of sales) are not presented in a meaningful context, are not combined with another measure, and there is no connection with a performance goal or best practice, which makes it operational data.

3d: Correct

3a,b,c,e: Incorrect

The operational data (the number of vacuum cleaners sold) are presented in a meaningful context (sold by a specific sales agent), a e combined with another measure (the past month), and is connected to a performance goal (target), but not to an industry standard, which makes it a KPI.

Did I get this? 8.5

d: Correct

a,b,c: Incorrect

Learn by doing 8.5

a: Incorrect > In the past year the total loss due to errors was 120 million US dollars, so the benefits of collecting the data will most likely outweigh the costs.

b: Correct > First, the evidence is most likely dispersed across the organization and stored in separate databases and/or information systems. Second, unless there is a common identifier attached to all these types of data, it will be very hard to link them to see whether there are meaningful relationships.

c: Incorrect > The executives should be able to get access to all organizational data, including sensitive data such as performance ratings.

d: Not quite right. In some organizations data such as supervisors' performance ratings and the error rate of a department may be considered politically sensitive, but in general senior executives should be able to obtain this information.

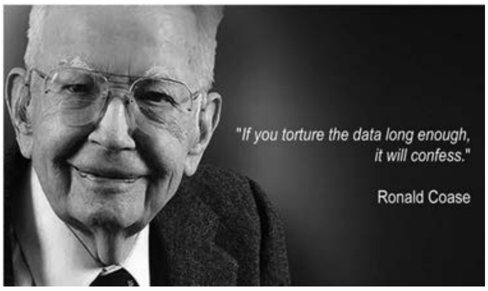
CEBMa
center for
Evidence-Based Management

Open Learning Initiative

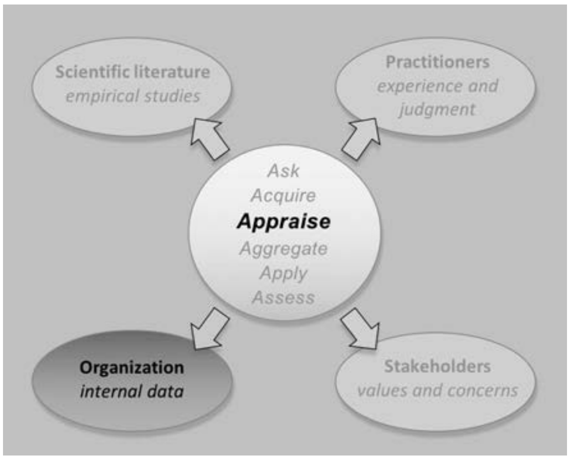
MODULE 9 | APPRAISE Evidence from the Organization

Learning objectives:

- Determine whether a logic model was used to collect and analyze evidence from the organization.
- Determine which organizational data are relevant.
- Determine whether organizational data are likely to be accurate.
- Determine whether contextual information is missing.
- Determine whether there could be measurement error.
- Determine whether a small number problem might exist, and if so, suggest remedies to solve it.
- Determine whether a metric is a good representation of the data.
- Determine whether a graph is a good representation of the underlying data.
- Determine whether a correlation or regression coefficient may be misleading.
- Determine whether a correlation or regression coefficient is practically relevant.
- Determine the precision of a finding or outcome.



Introduction



Scientists are well aware of the need to be skeptical. They acknowledge that research findings can be misleading, due both to methodological shortcomings and problems with the underlying data. As a result, even well-respected peer-reviewed journals sometimes publish flawed findings. In the realm of organizational evidence, the same fundamental problems arise, but practitioners are often less skeptical, unaware of the need to critically appraise their organizational evidence.

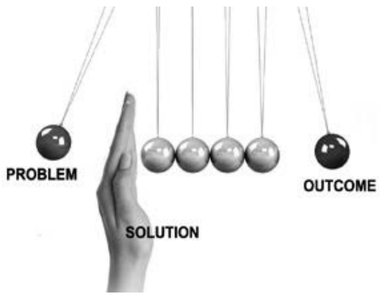
The decision-making process in many organizations resembles a form of competitive storytelling, where organizational evidence plays a small supporting role in the face of arguments practitioners use to persuade each other. This uncritical use of organizational evidence can lead to unnecessary – and unconscious – organizational risks.

A key point of this module is that managers not only need to ask *What do the data say?* but also need to get in the habit of asking critical follow-up questions, such as *Where did the data come from?*, *What kind of analyses were conducted?*, and *How confident should we be in the findings?* Hence, a fundamental step to improve the quality of decision-making is to learn how to gather relevant organizational evidence and critically appraise its quality. (1)

Before you start reading this module . . .

This module provides an overview of the most common barriers to overcome when using organizational evidence: data, facts, and figu es gathered from and about the organization. Gathering and analyzing organizational evidence is similar to gathering and analyzing scientific evidence. In fact, we argue that there is no fundamental difference between conducting scientific research and examining data gathered from the local organization. We must base both on the principles of the scientific method to ensure the trustworthiness of the findings. This means that most of what is discussed in modules 6 and 7 on the critical appraisal of scientific research also applies to the critical appraisal of organizational evidence. As a result, issues of bias, statistical significance, practical relevance, effect size, confounders, reversed causation, moderators, mediators, and the like should also be taken into account when evaluating the quality of organizational data. For this reason, we strongly recommend completing modules 6 and 7 before you start this module.

Barrier 1: Absence of a logic model



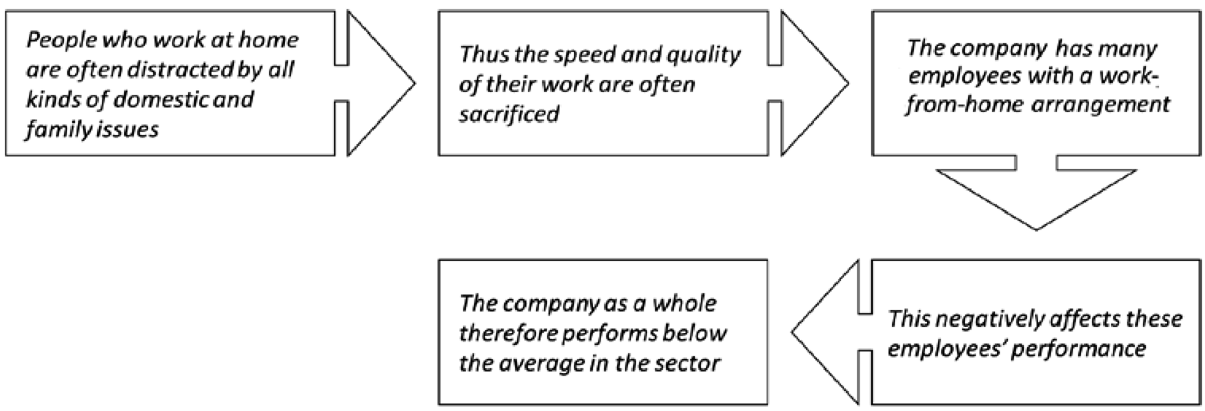
Using organizational evidence in the decision-making process is not about simply collecting and analyzing data to identify interesting patterns. An evidence-based approach starts by asking questions such as *What is the problem to be solved, and what is the evidence for this problem?* or *What is the preferred solution, and what is the evidence that this solution will be effective?* Organizational data are important to answering the second part of these two questions: *What is the evidence for either the problem or the solution?* To acquire evidence from the organization, we need to start by formulating a logic model.

A logic model (also called a *causal mechanism or theory of change*) spells out the process by which we expect an underlying cause to lead to a problem and produce certain organizational consequences. It is a graphic representation of the logical connections between inputs (resources, antecedents), activities and processes, and outputs and outcomes. Think of a logic model as a short narrative explaining why or when the problem occurs and how this leads to a particular outcome. Logic models provide a way of conceptualizing problems and processes in ways that incorporate more information and data points than is possible in unaided human judgment. Just as a day planner or agenda helps you keep in mind all the things to do in a day, a logic model helps you stay on top of the array of issues related to a particular problem, project, or situation.

Example

The CEO of a large IT company sends a memo to all of the company's 12,000 employees, stating that it is critical that all employees are present in their offices. "Some of the best work and ideas comes from working together, meeting new people, and impromptu team meetings. Speed and quality are often sacrificed when we work from home. To be competitive in the market we need to be top performers, and that starts with physically being together." She therefore cancels all work-from-home arrangements and instructs all employees to work in the office.

In this example, the logic model looks like this:



A logic model helps us determine which organizational data and information to collect and analyze. In the example above, organizational evidence should be obtained to answer the following questions:

1. How many employees frequently work from home?
2. How does this number compare to the total number of employees? Is this number substantial?
3. What is the average performance of employees who frequently work from home?
4. What is the average performance of comparable employees who do not work from home?
5. Is there a difference in performance between these two groups?
6. If so, is the difference practically relevant?
7. If so, what is the impact of this difference on the total performance of the company?

In this example, the issue was not only about performance, but also about 'new ideas' and innovations. However, organizational data and information about these two outcomes may be hard to obtain.

Did I get this 9.1

Read the following scenario.

During a meeting with the executive board, the company's sales director states: "I noticed that when newly hired sales agents fail to meet their sales targets over time, they are more likely to focus on immediate sales and disregard customer needs. As a result, their sales performance will decrease even further. I therefore propose to enroll them in a training program on adaptive selling techniques that teaches them to adjust their sales strategy in ways that better fit customers' needs and preferences – this would really increase their sales numbers and improve our financial performance."

In this example, how would the logic model look? Rank the six steps below in the right order.

A. Thus, their sales performance is likely to decrease.
B. As a result, their sales performance is likely to increase.
C. Disregarding customer needs and preferences could negatively affect sales performance.
D. A training program on adaptive selling techniques teaches agents to adjust their sales strategies to better fit customer needs and preferences.
E. When newly hired sales agents fail to meet their sales targets over time, they are more likely to focus on immediate sales and disregard customer needs.

Did I get this 9.2

Based on the logic model you formulated above, which organizational data and information would you collect and analyze? Select yes or no.

a) Whether the number of sales of agents who focus on immediate sales is lower than those who focus on customer needs
b) Whether newly hired sales agents who fail to meet their targets over time focus on immediate sales and disregard customer needs
c) The average number sales of the company's sales agents
d) The number of newly hired sales agents who fail to meet their sales targets for a longer period
e) The average number of sales of the company's new sales agents compared to the average in the sector
f) The costs of training all new sales agents

Finally, a logic model roots assumptions about problems (and preferred solutions) in “real” tangible relationships linked by evidence. Failure to use a logic model increases the chance of what is known as a "type I error": finding a relationship when none really exists. (2)

Example

Military intelligence analyst Tyler Vigen demonstrates various kinds of spurious, nonexisting correlations. (3) On his website, you can find examples of bizarre, nonexisting correlations, such as between the divorce rate in Maine and the per-capita consumption of margarine (r = .99) and between the marriage rate in Kentucky and the number of people who drowned after falling out of a fishing boat (r = .95). Similar strong but spurious correlations are found when large sets of organizational data are exhaustively analyzed, such as the nonsensical relationship between sales representatives’ shoe size and number of sales or the relationship between a firm’s financial performance and its CEO’s handicap with the US Golf Association. Without a logic model to guide our assessments and analysis, we are likely to identify other such meaningless relationships.

Learn by doing 9.1

Read the two scenarios below and determine whether the data were collected on the basis of a logic model.

Scenario 1

We have collected organizational data on absenteeism. After analyzing the data, we noticed that the absenteeism rate of employees born between 1981 and 1996 (also known as the millennial generation) is on average 20 percent higher than that of other employees. This suggests that millennials have a more lenient attitude toward sick leave. We should therefore focus our efforts to decrease the company’s absenteeism rate on this group rather than on all employees.

Scenario 2

Our employee turnover rate is 20 percent above the average in our sector. Recruiting, selecting, and on-boarding new employees costs a lot of money, so this is an opportunity to save costs. Because it is often said that employees don’t leave companies – they leave bosses, we analyzed 200 exit interviews. It turns out that several good employees did indeed leave because they had issues with their managers. We therefore analyzed the results of this year’s employee survey and found that 27 percent reported that their manager doesn’t give them credit for hard work, 19 percent feel their manager does not recognize their strengths and talents, and 23 percent feel that their manager is a poor communicator. These results suggest that some of our managers lack essential people skills. We therefore analyzed our managers’ educational backgrounds and found that most of them, especially at the supervisory level, never had any training in basic management skills.

Did I get this 9.3

Read the scenario below and determine whether the data were collected on the basis of a logic model.

As part of the company's Global Risk Management Program, we have analyzed the common characteristics of investment bankers who take excessive risks. After analyzing more than 300 critical incidents, several characteristics were conclusively indicated by a confirmatory factor analysis. Next, a one-way ANOVA tested whether these characteristics were related to the appetite bankers have for taking excessive risks. It was found that gender (male), age (between 25 and 35 years), marital status (divorced), communication skills (good), sense of humor (great), and car preference (BMW) have significant links to bankers' risk appetite. Thus, we recommend that activities of investment bankers who possess these characteristics be closely monitored.

Barrier 2: Irrelevant data

In most companies, organizational data are presented as a metric – an indicator of the organization's performance derived from a combination of two or more measures. Examples include financial ratios, daily occupancy rates, and average time per phone call. Metrics are an important input to management decisions if what they measure is meaningful (e.g., the turnover rate in a department's staff or the production volume of a manufacturing company). Unfortunately, it is all too easy with modern technology to create an array of colorful metrics, giving a false impression of understanding. For example, the number of sightings of unicorns by members of the staff at a food-processing company may provide insight into their mental health, but it is likely to be irrelevant to the company's financial performance

Metrics are likely to be more relevant when you assess outcomes rather than activities. For example, evaluating what sales agents do in terms of type of activity and time spent can provide a valuable source of data for investigating productivity, but an even more useful metric is the number of actual sales per agent per week. Unfortunately, organizations are often tempted to create metrics from easily available data rather than create metrics that add real value to management decisions. Keep in mind that more metrics do not always equate with better understanding. Organizations can be better off with a small number of relevant and accurate metrics than with a large number of fancy but irrelevant ones.

Don't collect organizational data just because it is easy to do so. Distinguish between what is easily measured and what is relevant to measure.(4) An important indicator of relevance is whether the data have a key place in your logic model.

Example

In an effort to drive sales, an automobile manufacturer launched a marketing campaign to draw more customers to its showrooms and thus increase car sales. The campaign encouraging free test drives seemed a big success, substantially increasing customer visits. Indeed, the metric indicating the number of visits per day went up by 20 percent. The number of sales, however, did not increase. Dealers offered the manufacturer an important explanation: "A lot more people are visiting the showroom, but most are 17- or 18-year-old boys, and they aren't going to be buying." This example illustrates how you may get a good result on a misleading metric unrelated to the desired outcome.

Other misleading or irrelevant metrics are those that are not actionable. For example, revenue is an important indicator of how a company is doing, but just knowing that revenue increased without knowing its possible cause is not very helpful: unless the metric indicates a significant change before and after a specific event or intervention, you lack essential information about how to sustain or amplify the result.

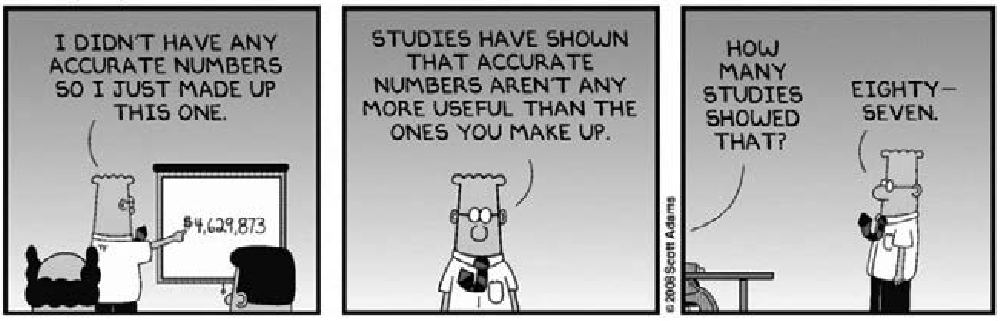
Did I get this 9.4

Read the following scenario.

A global food and beverage company hires a renowned advertising and marketing agency to develop an international sales campaign for a new candy bar. The agency develops a video ad that features a well-known comedian and a popular female movie star. In the first month after its launch, the video ad goes viral and is shared more than 5.4 million times across Facebook. A global survey of the candy bar's target group (young adults ages 16 to 24 years) shows that they find the video extremely funny, which is probably the reason the ad went viral. The food company regards the 5.4 million shares as an indicator of the campaign's success, and for this success, the marketing agency receives a big bonus.

Do you agree with the food company that the number of shares on Facebook is a relevant indicator for the sales campaign's success?

Barrier 3: Inaccurate data



Decades after the expression "garbage in, garbage out" was coined, many companies still struggle with data accuracy. It is important to remember that data aren't the same as facts. The metrics and key performance indicators (KPIs) presented in a management information system or business dashboard are numbers collected and recorded by *people*. These data *may* represent facts, but all kinds of biases and distortions can slip into them. By the time metrics and KPIs are reported to senior management, they can take the form of authoritative-looking numeric data. However, even subjective opinions can be transformed into numbers, and as a result, organizational data sometimes look more objective than they really are. An important question to ask is, *How were the data collected, processed, and reported?* In most companies, organizational data typically has gone through one or more of the following steps before being presented as metrics or KPIs.

Collection

Organizational data first need to be collected, also referred to as *data capture*. Ideally, a person or an automated system collects data routinely. As a rule, the collection of direct/objective data (e.g., production errors, staff turnover rate) is more likely to be accurate than that of indirect/subjective data based on self-report (e.g., self-reported accidents or perceived performance). In addition, the accuracy of self-reported data can be seriously reduced if the data are collected after the actual activity being described. People lack 100 percent recall, even after a short time delay.

Extraction

Ideally, data are extracted from the original source. The more steps the original source is from the final data capture system, the greater the risk of bias or inaccuracy. In addition, some methods of data extraction are more flawed than others. Copying data f om an Excel spreadsheet into another application by hand, for example, is less accurate than automatically extracting data directly into a database.

Aggregation

Organizational data may be captured and combined from multiple sources. Avoid combining organizational data of different types on the (invalid) assumption they are comparable (like comparing apples with oranges). For example, when a hospital collects data on patient treatments, it would make little sense to add together the treatment data of a department that deals with minor injuries with those of a department that performs open heart surgery. The data of these departments are too different in terms of impact, cost, time, and skills required.

Conversion and (re)formatting

When data are captured from different sources, they must be converted into a standard format. The conversion process may involve splitting data apart in some fields and combining data in other areas. Each stage of reformatting, however, introduces the risk of error and distortion of the original data.

Interpretation

Creation of metrics from raw data generally requires some form of interpretation. For example, KPIs often take the form of ratings on a scale. This requires defining a norm or threshold – for example, what data should be scored as low or high or as 1 or 2 or 3? Interpretation is typically carried out by people. If their analyses is not based on clear rules and guidelines, then KPIs are prone to bias and may be inaccurate.

Summarization

Large amounts of organizational data can be difficult to manage, let alone interpret, so it makes sense that metrics and KPIs are often a summary of the data. However, summarization typically requires some degree of subjective interpretation (see above) and may involve losing important information. While the steps described above are necessary to provide managers with an understandable set of metrics, the more steps the data go through, the more likely they will be inaccurate. As a rule, inaccuracy increases wherever people are involved in any of these steps – because of human error, incentives to manipulate the data, or even a simple lack of understanding of the data.

Did I get this 9.5

Read the scenario below.

A large European car insurance company has a toll-free 24-hour claims hotline that – based on area code – automatically routes customer calls to 26 local centers. The average number and duration of calls per center calls is presented in the management information system. The data are collected, processed, and reported according to the steps below.

Identify the two steps that introduce significant risk of inaccurate data

a) The number and duration of phone calls of each center is captured by an automated system on a daily basis.

b) The data are extracted into the company's central data system and automatically aggregated.

c) The average number and duration of calls of each center is automatically scored using a 3-point rating scale that is based on a norm that was set by the company's vice president of marketing.

d) For each center, the scores are summarized, interpreted, and assigned a RAG (red, orange or green) status by one of the company's five channel marketing managers

e) The RAG (red, orange or green) status of each center is automatically converted and presented in the company's management information system.

Barrier 4: Missing contextual information

Organizational data can be misleading without context. Consider the following statement made by a manager at a large bank: “My team is carrying out twenty-five activities that other teams should be doing instead!” The number 25 was repeated meeting after meeting until a project team was formed to migrate the activities. After one year’s effort, there had been little progress, because the activities were generally too complex and expensive to move. An additional question was then asked about the 25 activities: “How long do those activities take to complete in a typical month?” To everybody’s surprise, it turned out that most of the activities did not take place very often and on average only a few hours per month was spent on all of them combined.

Metrics and KPIs empty of important contextual information are surprisingly common. In the example above, looking only at the number of activities without considering total time spent leads to a considerable amount of money being spent on a project of little value. The same applies to organizations that use “headcount reduction” as a metric for cost savings. For example, a cardiologist’s average base salary in the United States is $550,000 (5), whereas the annual salary of a nursing aide is only $25,000 (6), so a headcount reduction of 10 percent is a meaningless metric when information about the functions involved is missing.

Other contextual information often missing is the average in the sector, setting, or industry. A staff turnover rate of 200 percent annually sounds pretty terrible until we learn that the organization is comprised of fast food restaurants and the industry average is over 300 percent. Widely used metrics such as absenteeism, staff turnover, job satisfaction, failure frequency, customer complaints, and so on, are meaningless when we can’t compare them to the average in the sector.

Finally, another contextual factor that is often overlooked is size. Whenever organizational data in the form of a metric is presented, always ask yourself, “Is that a big number?” (7)

Learn by doing 9.2

1. **The annual records of a large multinational corporation show that, over the past 5 years, the company has spent $2 million to provide subsidized childcare for 20,000 employees.**

   Would you consider $2 million to be a big number?

   a) No
   b) Can’t tell
   c) Yes

2. **The organizational data of the back office of a health insurance company indicates that more than 1,500 errors are being made every year, accounting for an annual loss of $1.5 million.**

   Would you consider the annual loss of $1.5 million to be a serious problem?

   a) Yes
   b) Can’t tell
   c) No

Did I Get This 9.6

In the past month, Bank X has lost 100 clients, whereas Bank Y has lost more than 1,000 clients. What contextual information would you need to determine which of the two banks is better off?

Barrier 5: Measurement error

Most organizational data can be considered as measurements: estimations of quantities such as an amount, volume, or frequency of something. A measurement is typically expressed as a number of standardized units, such as products sold, errors made, or employees hired. Unfortunately, whenever something is measured, its score is likely to deviate somewhat from its true score. This deviation is referred to as *measurement* error.

Of course, we can measure some things more precisely than others. For example, the measurement of a company's total revenue may be more precise than the measurement of its corporate reputation or employee morale. In addition, if outcomes are measured by people using different measurement methods, they will probably get different results.

Learn by doing 9.3

Read the following scenario.

A large US Midwestern manufacturing organization has experienced multiple restructurings and downsizings, consequently reducing its workforce from more than 8,000 to fewer than 6,000 employees. The restructuring and downsizings have been very stressful for employees and have led to anxiety and fear of job loss. The company therefore decided to deploy a stress-reduction program that includes on-site chair massage therapy, a technique that has successfully been tried by several companies, including AT&T, Apple, and Google. To determine whether the program is effective, it is decided to measure the participants' stress level both before and after the program's start.

Rank order the four measurement methods based on the chance of measurement error (the most precise method on top).

A. A third person observes whether a participant is stressed.

B. The participant fills out a validated questionnaire that is used in studies to measure stress.

C. A third person takes a sample of the participant's saliva, which is then sent to a laboratory to analyze the level of cortisol (a reliable biomarker for stress).

D. The participant indicates on a five-point scale how stressed he or she is.

Learn by doing 9.4

Let's say you were measuring the weights of 10 managers. Let's also say your scale was 100% precise. What reasons could there be for measurement errors still to occur?

Even if we measure something perfectly, there always will be a measurement error. This is not necessarily a problem, especially when the measurement error is negligibly small. However, small measurement errors tend to be larger when the metric is a *difference score*: a score that is derived by subtracting one variable from another. A widely used difference score is profit: sales minus costs. As a rule, difference scores have a large measurement error (and thus a lower reliability) when the two variables composing it are correlated. In other words, two variables (e.g., sales and costs) measured with very little error can yield a difference score (e.g., profit) with more measurement error.

Example

The organizational data of a global entertainment company shows that the annual profit over 2015 was $986 million. This number was calculated by subtracting the annual costs from the total sales. Both sales and costs were measured quite reliably. Sales and costs, however, are highly correlated. Thus, although little measurement error exists in the sales and costs data, there may be a large measurement error in the profit metric. As a result, the reported profit could be substantially lower (or higher) than the actual profit. (8)

Learn by doing 9.5

Which of the following metrics and KPIs are difference scores (and are thus more likely to deviate from the true score)? Check all three that apply.

a) Gross margin = (total sales revenue) minus (cost of all products sold) divided by (total sales revenue)

b) Customer turnover rate = (number of customers who left) divided by (total number of customers

c) Training expenses per employee = (total expenses) divided by (number of employees)

d) Return on investment (ROI) = (gain from investment) minus (cost of investment) divided by (cost of investment)

e) Customer growth rate = (number of new customers) minus (number of customers who left) divided by (total number of customers)

f) Average employee engagement score = (added up individual engagement scores) divided by (total number of employees)

You should *always* be aware of measurement error in using organizational evidence. Error can seriously distort metrics and KPIs. Sound methods exist to adjust for measurement error, getting you closer to the true estimate.

Barrier 6: The small number problem

A notorious problem with organizational evidence is the small number problem. This problem stems from what is known in probability theory as the *law of large numbers*. This law states that the larger the sample size, the more accurate its predictions regarding characteristics of the whole population, and thus the less the sample value deviates on average from the population value. As a general principle, the average (mean) of a large sample of organizational data will be closer to the *true* average of the total data than the average drawn from a small sample. Thus, when we use a small sample of organizational data, the metrics and KPIs based on that sample are most likely to deviate from the true value. The small number problem often arises in three situations:

1. **When organizations compare units (e.g., teams, departments, divisions) unequal in size.**
   Smaller units are more likely than larger units to report data that deviate from the true value. For example, when the monthly average absentee rate of the total number of employees is 2.5 percent, a small unit is more likely than a large unit to yield an average with a greater deviation from this 2.5 percent. This deviation has nothing to do with the employees' health status or the quality of its management but results solely from the small number problem. The same is true, of course, for metrics such as daily occupancy rate, number of defects per 1,000 products, or average handle time per phone call. Metrics from smaller units will deviate more than those from larger units.
2. **When organizations collect data from a sample rather than from the whole organization.**
   This makes sense, as collecting data from the whole organization is often time consuming and thus expensive. However, the smaller the sample, the greater the chance that the values measured deviate from the true value. For example, if only a small sample of employees participates in a survey, the data derived from this sample will most likely deviate from the data derived from the total population.
3. **When organizations have access only to a small sample of the total market population.**
   For example, insights into customer behavior based on the data drawn from an organization with a small market share are less likely to be representative of the whole market population than are insights based on the data of customers of a global brand such as Nike, L'Oréal, or Heinz.

Learn by doing 9.6

1. **Read the following scenario.**

A town is served by two hospitals. In the larger hospital, about 45 babies are born each day, and in the smaller hospital, about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent and sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys.

Which hospital do you think recorded more such days?

a) The larger hospital

b) The smaller hospital

c) About the same (within 5% of each other)

2. **Read the following scenario.**

   Business unit A has 10 employees and 20 percent turnover; business unit B has 55 employees and 10 percent turnover.

   Is employee retention better in business unit B?

   a) Yes
   b) No
   c) Can't tell

Did I get this 9.7

Read the scenario below.

You are responsible for the management information system of a large retail company that operates a chain of discount department stores. The company has about 2,300 stores. The average absenteeism rate in the industry is about 3 %. However, the management information system shows that at 65 stores the absenteeism rate is above 6%. The company's HR director asks you how he should interpret these numbers.

What would be your answer?

a) The absenteeism rate at the 65 stores is too high
b) The absenteeism rate at the 65 stores is NOT too high.
c) It is unclear whether the absenteeism rate at the 65 stores is too high.

A key approach for dealing with the small number problem is simply to be aware of it. In addition, the following four remedies are helpful:

1. Sample from the whole population.
2. Aggregate internal data to achieve larger sample sizes.
3. Pool data from several firms (this occurs in many industries) to develop a larger data set
4. Clearly state the sample size and report confidence intervals (explained later in this module)

Barrier 7: Confusing percentages and averages

Confusing percentages

We use percentages to express a proportion out of a total – for example, "67 percent of all employees are male," or "the failure rate of a newly designed product is 0.5 percent." We also use percentages to express differences or degree of change – for example, "the absentee rate in the company has gone up by 5 percent."

Change and differences, however, can be relative or absolute. An absolute change (increase or decrease) expresses the difference between two comparable values, such as differences in products sold, total revenue, or employees hired at two points in time. For example, if last year's revenue was $5 million and this year's revenue is $4 million, then this is an absolute difference of $1 million.

Absolute change or differences are useful when the amount of change itself is relevant, regardless of the base value. For example, an increase of $1 million in revenue may be relevant regardless of whether that increase came from a large company with an annual financial turnover of $100 million or a small company making only $5 million a year. Relative change, on the other hand, expresses difference as a percentage of the base value. In the example above, for the $5 million company, an increase of $1 million is 20 percent, whereas for the $100 million company, it is an increase of only 1 percent. Although the absolute change is the same for both companies ($1 million), the relative change is very different (1 percent versus 20 percent).

For this reason, it is often assumed that relative change expressed as a percentage of the base value is a more accurate metric. As we demonstrate in the example below, however, this is not always the case.

Learn by doing 9.7

Read the following scenario.

A pharmaceutical company has tested a new experimental drug for Parkinson's disease. Compared with drugs currently prescribed, the new drug decreases symptoms such as tremors, limb stiffness, impaired balance, and slow movement by 30 percent. However, compared with the existing drugs, the mortality rate of patients taking the new drug (those dying because of serious side effects) has increased by 200 percent.

Would you decide to bring this new drug to the market?

a) Yes

b) No

c) I need more information.

Did I get this 9.8

Read the following scenario.

In a large Canadian hospital, the number of medication errors is measured on an annual basis. In 2016, unit A treated a total of 50,000 patients of whom 526 received the wrong medication. In 2017, the number who received the wrong medication was 1,050, an increase of 100 percent. In unit B, this increase was only 15 percent.

Is patient safety worse in unit A?

a) Yes

b) No

c) Their patient safety is about the same.

d) I need more information.

Whenever changes or differences are presented as percentages, we must make clear whether these differences are relative or absolute. Ideally, both types – including the number of standardized units they represent – should be reported.

Finally, determine whether the data concern *percent* or *percentage point*. If the unemployment rate last year was 4.8 percent and this year it is 6.0 percent, is that an increase of 25 percent? (6.0 – 4.8 = 1.2 and 1.2 ÷ 4.8 = 0.25, which equals 25 percent). Or is it just an increase of 6.0 – 4.8 = 1.2 percent? We can use either, but to avoid confusion, the latter is referred to as a percentage point. In this case, the government would probably use 1.2 percentage points, whereas the opposition would prefer 25 percent. There is no rule here, so always ask the underlying data. (9)

Confusing averages

Much of the organizational information used by companies is expressed as averages. Like percentages, averages look simple, but that simplicity is deceptive. In fact, there are three ways to calculate an average, and each yields a different number. For this reason, we avoid the term average and use instead the more precise terms mean, median, and mode. In some cases, they are identical, but often they are not. In addition, when the word average is used, this usually refers to the mean, but unfortunately not always.

Example

Five people are sitting in a restaurant. Each earns about $100,000 per year. Here are their earnings:

Person 1:	$96,000
Person 2:	$96,000
Person 3:	$99,000
Person 4:	$104,000
Person 5:	$105,000

The mean is calculated by adding all observations (e.g., reports, metrics) and then dividing the outcome by the number of observations. Here the mean is exactly $100,000. The median is the middle number in a set of numbers. In this case, the median is $99,000. The mode is the most frequent number in a set. Here the mode is $96,000.

In the example above, the mean, median, and mode are more or less of the same value. This, however, is not always the case. For example, let's assume Bill Gates walks into the restaurant (and his annual earnings are assumed to be $1 billion). The mean in our restaurant example has increased from $100,000 to more than $166 million. But the point of an average is to represent a whole pile of data with a single number. In the first example, the mean is a good representation of these five people s annual incomes, but in the second example, when Bill Gates shows up, it is not, so we should use the mode or the median instead.

Learn by doing 9.8

A pharmaceutical company's management information system shows the following number of weekly sales by their 18 sales agents:

40, 32, 40, 50, 30, 40, 35, 37, 34, 45, 44, 31, 40, 42, 46, 40, 39, 55

Which average represents the data better: the mean, the median, or the mode?

a) The mean
b) The median
c) The mode
d) The mean, media, and mode are equally representative.

Did I get this 9.9

Six months later, after hiring some new sales agents, the sales numbers have changed:

31, 43, 42, 47, 41, 60, 61, 60, 70, 39, 70, 71, 72, 72, 73, 72, 72, 72, 72

1. **Which average represents the data better: the mean, the median, or the mode?**
   a) The mean
   b) The median
   c) The mode
   d) The mean, media, and mode are equally representative.

Finally, after a year, when several experienced sales agents left the company, the sales numbers have changed again:

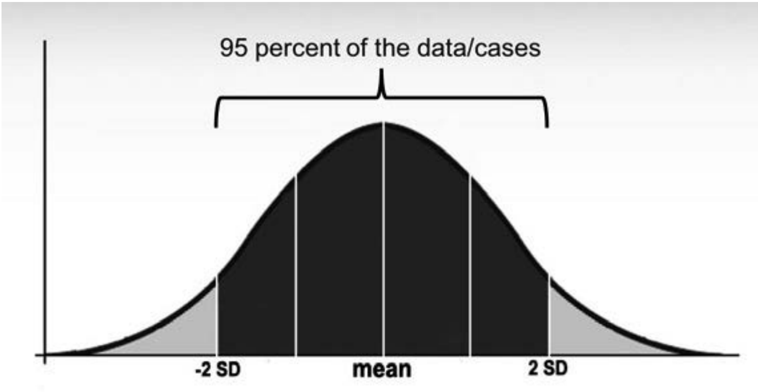
39, 42, 42, 40, 44, 45, 39, 43, 142, 182, 132, 41, 44, 44, 44, 44, 44, 44, 45

2. **Which average represents the data better: the mean, the median, or the mode?**
   a) The mean
   b) The median
   c) The mode
   d) The mean, media, and mode are equally representative.

An average is, by definition, an inaccurate metric: it combines a large amount of data and represents it with a single number. As a result, averages tend to obscure the natural variance. Thus, whenever a metric represents an average, you should look for an indication of what the variance is, that is, how the numbers are spread out around the average.

Standard Deviation

To determine whether an average is a good metric to represent the data, statisticians have come up with the *standard deviation*, a measure that tells us how much the data deviates from the average. For example, when the mean age of all employees is 40 and the "standard deviation" is 10, this tells us that the typical deviation from the mean is around 10 years. Now, here is a very useful rule of thumb: about 95 percent of the data will be covered by *two standard deviations* plus or minus the average.



Thus, in the example above, a standard deviation of 10 tells us that 95 percent of the employees are between 20 years (40 – [2 × 10]) and 60 years old (40 + [2 × 10]). Likewise, when the mean age of all employees is 50 years and the standard deviation is 5, then 95 percent are between 40 and 60 years old, indicating that the company's workforce is rather senior.

Learn by doing 9.9

The mean age of a company's workforce is 35, with a standard deviation of 5. What does this tell you?

a) About 95 percent of the company's workforce is between 20 and 55 years old.
b) The company's workforce is rather old.
c) About 95 percent of the company's workforce is between 25 and 45 years old.
d) About 95 percent of the company's workforce is between 30 and 40 years old.

Did I get this 9.10

A company's absenteeism rate is 10 percent with a standard deviation of 5. The average rate in the sector is 15 percent. What does this information tell you? Check the two statements that apply.

a) The metric of 10 percent is not a good representation of the company's true absenteeism rate.
b) It is unclear whether the company's absenteeism rate is on par with the average in the sector.
c) The company's absenteeism rate is on par with the average in the sector.
d) The metric of 10 percent is a good representation of the company's true absenteeism rate.

The standard deviation (often abbreviated SD) is also helpful in determining the size of a change or difference. If you take the absolute change and divide it by the standard deviation, you get a good impression of its magnitude. In the social sciences, a change of 0.2 is usually considered a small difference, while 0.5 is considered a moderate difference, and 0.8 is a large difference. (10)

Example

In the past 4 years, an Italian shoe factory has experienced multiple restructurings and downsizings, reducing its workforce from 800 to fewer than 500 factory workers. The HR director believes that this has been very stressful for the workers, causing a dramatic productivity decline. He decides to introduce "Search Inside Yourself," a mindfulness-based stress-reduction program that is successfully used at AT&T, Apple, and Google. A few months after the program is introduced, organizational data indicate productivity has gone up: the workers' average (mean) productivity has increased by 6 percent from 200 shoes to 212 shoes per day, with a standard deviation of 7. An absolute change (increase) of 12 shoes equals 12/7 = 1.7 This suggests that the program may have had a large impact.

Learn by doing 9.10

Read the following scenario.

The large Canadian hospital that we used earlier as an example decides to implement a program to improve patient safety. After 6 months, the average (mean) number of medication errors has decreased by 10 percent from 55 errors to 49.5 errors per month, with a standard deviation of 20 errors.

How would you judge the size of this decrease?

a) Small

b) Moderate

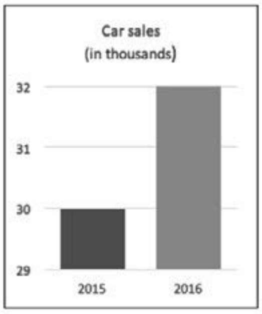
c) Large

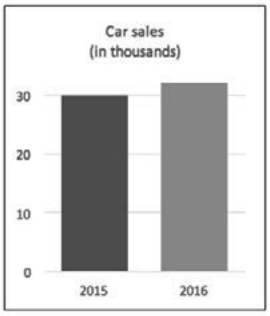
d) I need more information.

Barrier 8: Misleading graphs

Many specialized tools are available for presenting organizational data, such as management information or business intelligence systems. Most include advanced dashboards and visualization tools that present data in graphs, charts, and appealing visualizations. When used appropriately, graphs such as pie charts, bar graphs, and scatterplots help us to intuitively grasp complex data. Graphs, however, can be misleading. Following are some well-known examples.

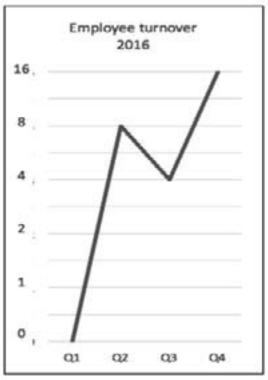
Example 1: Omitting the baseline

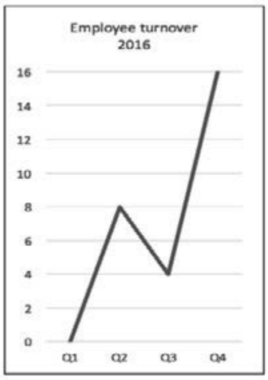




From the graph on the left, it looks as if the number of car sales has tripled in one year. This graph, however, is misleading because the baseline is missing: the vertical axis does not start at 0. The right graph presents the same data, but here the vertical axis correctly starts at 0. This graph indicates that the increase in sales was rather small.

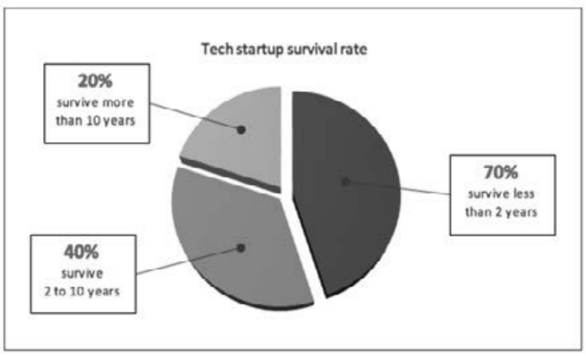
Example 2: Manipulated axis





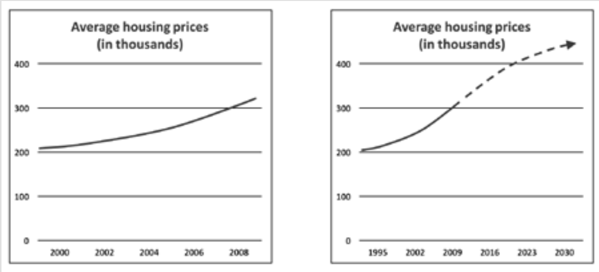
Although the vertical scale in the left graph correctly starts at the baseline (0), it does not go up in equal steps. This distorts the graph (as you can see in the right graph using equal steps) and makes it look as though the highest turnover was in Q1 and Q2 rather than Q3 and Q4. In addition, it is not clear what the numbers on the vertical scale (0 to 16) mean.

Example 3: The numbers don't add up



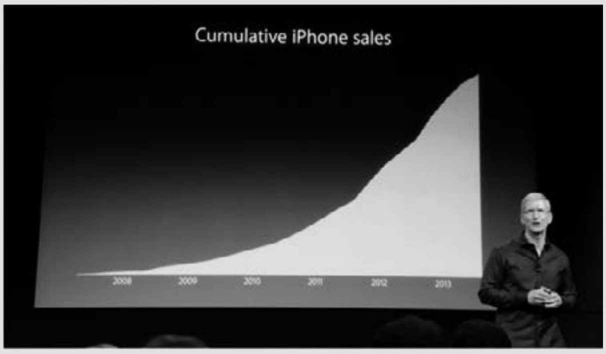
This pie chart measures the survival rate of technology startups. Note its categories are mutually exclusive and its percentages should add up to 100 percent, not 130 percent.

Example 4: Cherry-picking



The graph on the left shows that between 2000 and 2008, the average housing prices have been steadily growing from 200,000 to 300,000. As you can see in the graph on the right, this increase can be visualized more dramatically by extending the X-axis into those years when no data are available. As a result, the slope of the curve increases, suggesting that the prices have gone up faster than they really have. In addition, a dotted trend line is added, suggesting that prices will keep growing for the coming decades. Note that the dotted line starts at 2009, one year after the housing market collapsed.

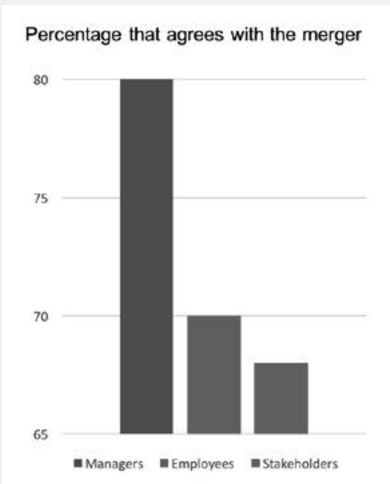
Example 5: Cumulative versus interval data



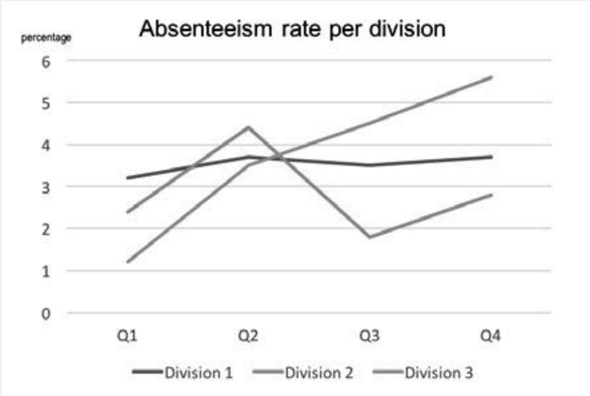
Instead of plotting data for separate time intervals, sometimes cumulative data are presented. For example, the number of sales can be visualized for each month, quarter, or year separately, but it is also possible to present them as cumulative sales per month, quarter, or year – that is, total sales to date. When the cumulative sales are presented in a graph, each single sale makes the graph's line go up. This is particularly convenient when the total number of sales is decreasing, and it is exactly what Apple's CEO Tim Cook did during a presentation on iPhone sales in 2013. (11) In the picture above, you can see that the graph suggests that the iPhone sales are still increasing. However, by plotting cumulative sales instead of quarterly sales, Apple hid the fact that iPhone sales were actually declining. (12) If you look carefully at the graph, you can see that after 2013, the line is going up less steeply, indicating a decrease in sales.

Did I get this 9.11

Look at the graphs below and determine whether (and why) they are misleading.



a) Misleading: the baseline is missing – the vertical axis does not start at 0.
b) Misleading: the vertical axis is distorted – the steps are unequal.
c) Not misleading: the graph is correct.
d) Misleading: the graph presents cumulative rather than interval data.



a) Not misleading: the graph is correct.
b) Misleading: the baseline is missing.
c) Misleading: the numbers don't add up.
d) Misleading: these are cumulative data.

Barrier 9: Correlations and regressions

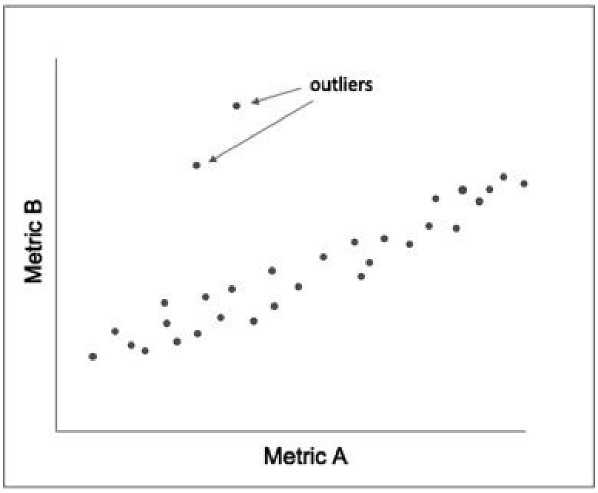
In Module 6, A Short Introduction to Science, we explained that two metrics are correlated when a change in the value of one metric leads to a change in the value of the other. For example, temperature and ice cream sales are correlated; when the temperature increases, ice cream sales go up as well. Regression involves the prediction of an outcome metric from one predictor metric (simple regression) or several predictor metrics (multiple regression). For example, for every one-degree rise in temperature, about 1,200 more ice creams are sold on average.

Correlations and regressions are ***inferential*** measures. Whereas descriptive measures such as operational data and KPIs provide basic information, correlations and regressions allow us to infer trends, identify relationships, and make predictions about organizational outcomes. However, correlations and regressions can be misleading and therefore require the same scrutiny as percentages and averages.

Correlation coefficients

A correlation coefficient is a numerical index that reflects the strength of the relationship between two variables. There are different types of correlation coefficients, depending on the type of variables we a e measuring, but they all have a value that ranges between −1 and +1. A correlation of r = 0.10 is considered a weak relationship, whereas a correlation of r = 0.60 is regarded as a strong one. A correlation coefficient can be influenced by several factors, including ***outliers***, that is, a value or data point that differs substantially from the rest of the data. In the example used in the section on averages, Bill Gates's annual earnings of $1 billion is an outlier, as the other five people earn much less than this, about $100,000. Therefore, you should always check the ***scatter*** plot, a graph that shows the relationship between two metrics and helps identify outliers.

r-Squared: Variance explained



To get a better idea of how strongly two metrics are correlated, we can take a look at the r^2 (pronounced "r-squared"). The r^2 indicates the extent to which variation or differences in one metric can be explained by a variation or differences in a second metric. The r^2 is expressed as a percentage and can easily be calculated by squaring the correlation coefficient. For example, if the organizational data show that the correlation between customer satisfaction and sales performance is r = 0.5, then the r^2 is 0.25, indicating that 25 percent of the variation in sales (increase or decrease) can be explained by a difference in customer satisfaction.

Did I get this 9.12

Read the following scenario.

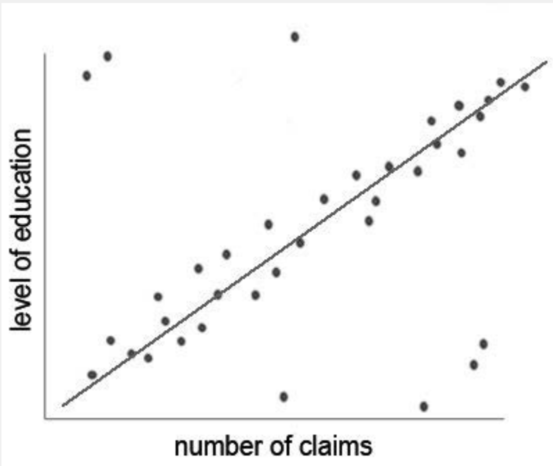
A car insurance company's data analytics system indicates that there is a statistically significant correlation between a client's weight and the likelihood that a client will file a claim. The correlation found is r = 0.12. The company's sales director therefore recommends that clients with a body mass index (BMI) that fall in the obese range (BMI > 30) pay 20 percent more insurance premium.

Given the correlation of 0.12, would you support this recommendation? Why or why not?

a) Yes, because the correlation is statistically significant
b) No, because the correlation may be nonexistent.
c) No, because the correlation is practically irrelevant.
d) Yes, because the correlation is practically relevant.

Did I get this 9.13

What would you conclude on the basis of the scatter plot?



a) There are outliers that may affect the correlation between education and number of claims.
b) There are outliers, but they don't affect the correlation between education and number of claims.
c) The correlation between education and number of claims may be spurious or nonexistent.
d) There is no practically relevant correlation between education and number of claims.

Range restriction

Another aspect to consider when judging correlations is ***range restriction***. Range restriction occurs when a metric in a data set has a more limited range (minimum and/or maximum value) than it has in the whole population. As a result, the correlation between that metric and another metric can be constrained.

Example

A midwestern branch of an American trading company analyzed its organizational data to see whether some factors strongly correlated with the performance of sales agents. It gathered the data of the top 20 percent of sales agents based on their monthly sales performance. The assumption was that both age and years of experience were important factors. When they analyzed the data, however, they found only small correlations. After some months, it became clear that the company's northeastern branch had conducted the same analysis but found that the correlation with sales performance was 0.36 for age and 0.55 for years of experience. When the data analysts of the two branches examined this remarkable *difference, they discovered that the northeastern branch had included the data of all sales agents rather than only the top 20 percent. It was therefore concluded that the low correlations of the midwestern branch were due to range restriction. This means that the midwestern branch used only one part of the sales data (the high end of the distribution), reducing the possibility of observing a relationship between sales agent characteristics and their performance.*

Range restriction can occur because the dataset comprises a subset of the total data (as in the example above) but also because the organization itself is a subset.

Learn by doing 9.11

In the general population, the correlation between general mental ability (IQ) and performance is about 0.6. In a prestigious law firm that recruits only lawyers with a degree from Stanford, however, this correlation will most likely be closer to zero.

Explain why.

Regression coefficients

A *regression coefficient* tells you how much the outcome metric is expected to increase (if the coefficient is positive) or decrease (if the coefficient is negative) when the predictor metric increases by one unit. There are two types of regression coefficients: unstandardized and standardized. An unstandardized coefficient concerns predictor and outcome metrics that represent "real" units (e.g., sales per month, points on a job satisfaction scale, numbers of errors). In that case, the coefficient is noted as *b*. For example, when a predictor metric temperature is regressed on the number of ice creams sold, a regression coefficient of $b = 8.3$ means that for every 1 degree rise in temperature, 8.3 more ice cream bars are sold on average.

A standardized coefficient involves predictor and outcome metrics expressed in standard deviations. In that case, the coefficient is noted a β (beta). Betas provide information about the effect of the predictor metric on the outcome metric. As explained in Module 6, in a simple regression, a β of 0.10 is considered a small effect, whereas a β of 0.60 is considered a large effect. In a multiple regression, the thresholds are slightly higher ($\beta = 0.20$ is considered small; $\beta = 0.80$ is considered large). Both *b* and β provide important information: *b* tells us exactly how much the predictor metric is expected to change, whereas β (which expresses that change in standard units) informs us whether that change is considered small, moderate, or large. For this reason, both unstandardized (*b*) and standardized (β) regression coefficients should be reported.

Did I get this 9.14

In the example of the car insurance company, the data analyst decides to regress the level education (predictor metric) on the number of claims. The analyst finds a regression coefficient of b = −2.1. Note that the coefficient is negative.

What does this mean?

a) This means that for every (one) higher level of education, on average 2.1 fewer claims are being made.

b) This means that for every (one) higher level of education, on average 2.1 more claims are being made.

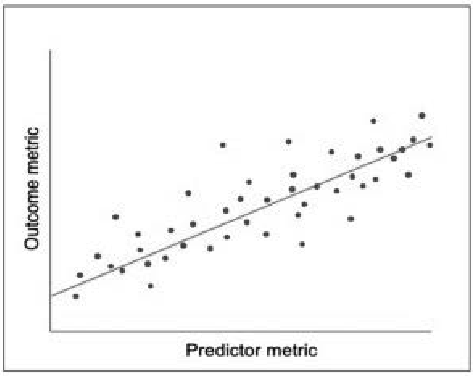
c) This means that the level of education does not predict the number of claims.

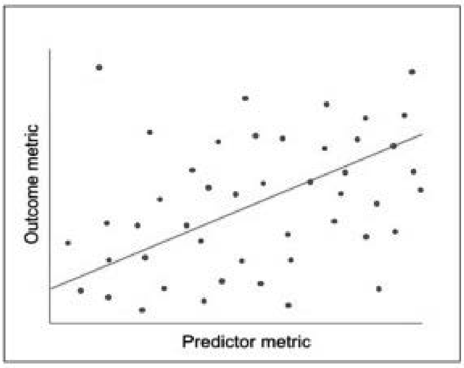
d) This means that the number of claims predicts the level of education.

Multiple regression tells you how the outcome metric is expected to increase (or decrease) when several predictors are considered at the same time. The logic used to determine the factors that contribute to a particular outcome metric can provide insight into which predictors to use. For example, a logic model specifying how new car sales are driven by (1) advertising to appropriate market segments (e.g., drivers who are not teenage boys), (2) sales agent skills, and (c) other predictors (e.g., discounts, special offers) can be used to run a multiple regression analysis to test whether its predictions are supported.

Multiple r-squared: Goodness of fit

Several factors should be examined when critically appraising a regression, such as variable types, multicollinearity, distribution, and linearity. Unfortunately, it is not possible to address all these aspects in this module. (13) In addition, it is always recommended to check the ***residual plot***, a graph that shows if (and how) the observed (true) values in a data set differ from the values as predicted by the regression. This is also referred to as *goodness of fit*. Following are two examples of a residual plot. The blue dots represent the observed values, and the line represents the values as predicted by the regression equation. As you can see, the observed values in the left plot are closer to the regression line than those in the right plot, indicating a better goodness of fit





Unfortunately, a residual plot is often not available. In that case, we can use the multiple *R*-squared (R^2) instead. (Note that it is a different indicator than the r^2 discussed above, which involves squaring a correlation coefficient to find out how much covariation exists between two metrics. That is why we use a lower case *r* for the variance explained and a capital *R* for the goodness of fit). In a regression analysis, the multiple R^2 tells us how close the observed data are to the regression line. Put differently, it is the percentage of the outcome metric that, based on the regression coefficient, is predicted by the predictor metric. For example, when the unstandardized regression coefficient b for customer satisfaction and the number of sales is 30.2, it indicates that for one point of improvement in the level of customer satisfaction, on average 30.2 more products are sold. However, when the multiple R^2 is only 0.18, it means that the level of customer satisfaction can predict only 18 percent of the number of sales.

Did I get this 9.15

In the example of the car insurance company, the data analyst found a regression coefficient of b = −2.1, meaning that for every (one) higher level of education, on average 2.1 less claims are being made. The data analyst, however, points out that the goodness of fit (multiple R^2) is 0.5.

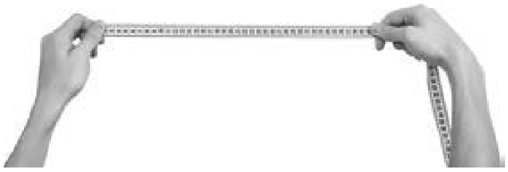
What does this mean?

a) This means that for every (one) higher level of education, on average 0.5 × 2.1 = 1.05 fewer claims are being made.

b) This means that a client's level of education predicts 50 percent of the number of claims.

c) This means that a client's level of education predicts only 5 percent of the number of claims.

Barrier 10: Wide confidence interval



In Module 6, A Short Introduction to Science, we explained that effect sizes such as correlation and regression coefficients di fer across samples, meaning that they are only an estimate of the "true" effect size. The same is true of estimates such as percentages, averages, and the R^2. This means that whenever a sample or subgroup of the total population (or whole data set) is used, we need to know how precise the estimate is. We can determine the precision of a percentage, mean, b, β, R2, or any other estimate by looking at its confidence interval (CI). A confidence interval provides the upper and lower boundaries between which we expect – usually with a 95 percent confidence – the true value to fall. A confidence interval is stated as "95% CI." If the 95% CI is fairly narrow, then the estimated value is a precise reflection of the "true" value. A 95% CI is considered too wide if the decision you would make based on the value of the lower boundary of the interval would be different from the decision you would make based on the value of the upper one.

Example

A British car insurance company routinely measures the satisfaction of its clients. In the last annual quarter, the level of customer satisfaction dropped from 7.5 to 7.0 on a 10-point scale. Because customer satisfaction has shown to be a good predictor for the number of sales (R_2= 0.62), the sales director feels that action should be taken. The HR director, however, points out that the customer satisfaction metric is based only on a sample of the company's total number of clients, so he asks the company's data analyst to calculate a confidence interval. This 95% CI turns out to be between 6.2 and 7.8, meaning that there is a 95 percent confidence that the true level of customer satisfaction falls somewhere between these two values. Both the sales director and the HR director consider this confidence interval to be too wide and therefore decide to take no further action at this point. In effect, it is possible that customer satisfaction remains at least as high as in the last quarter.

As the example above illustrates, it is important that whenever a percentage, average, coefficient, or any estimate is provided, you should always check (or ask for) the confidence interval

Learn by doing 9.12

Read the following scenario.

A Japanese car manufacturing company routinely measures the number of defects in the cars it produces. In the last annual quarter, the number of defects per 1,000 cars increased from 2.5 to 5.7. The number of defects, however, is a point estimate based on a sample of the total number of cars the company produces. For this reason, a confidence interval is calculated. This 95% CI turns out to be between 5.3 and 6.1.

How would you assess this confidence interval?

a) The 95% CI is too wide.
b) The 95% CI is sufficiently precise.
c) The 95% CI is too small.
d) I need more information.

Did I get this 9.16

Read the following scenario.

A Dutch academic hospital measures the absenteeism rate of its medical departments and related institutes on a monthly basis. In the last three months, the absenteeism rate of the Anaesthesiology & Pain Management Institute increased from 3.5 to 5.2. The absenteeism rate, however, is an average (mean) based on a sample of the total number of nurses and physicians employed at the institute. For this reason, the absenteeism rate accompanied by a confidence interval: CI 95% = (3.2–7.2).

Based on this confidence interval, would you consider taking further action to lower the institute's absenteeism rate?

a) No, because the CI 95% is too wide.

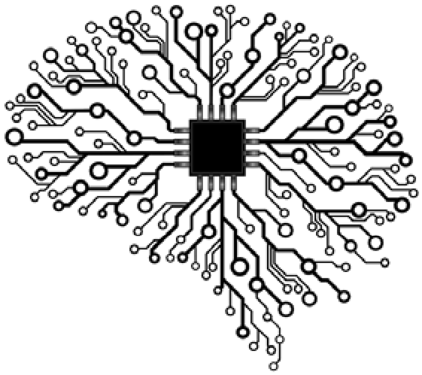
b) Yes, because the CI 95% is sufficiently precise.

c) I need more information.

Overview of appraisal questions

	Yes	No	Unclear
1. **Is the collection of organizational data based on a logic model?**			
2. **Are the data relevant to the organization's decision-making processes?**			
3. **Are the data accurate?** (Consider: How were the data captured? How were the data processed? Were the interpretation and summary of the data based on clear rules and guidelines? How often were people involved in these steps?)			
4. **Are the data's context taken into account?**			
5. **How reliable are the data – could there be a measurement error?** (Consider: Are the data based on direct/objective outcome measures or indirect/subjective measures? Are the reported measures difference scores?)			
6. **Is the data set sufficiently large enough to prevent the small number problem from occurring?**			
7. **When a change or difference in the form of a percentage is presented, is it clear whether this involves a relative change or an absolute change?**			
8. **When an average (mean, median, mode) is presented, is it clear what the variance is – is the standard deviation reported?**			
9. **When a graph is presented, could it be misleading?** (Consider: Is there a baseline? Do the units on the axes represent equal steps? Do the numbers add up? Does the graph present missing data? Does the graph present cumulative or interval data?)			
10. **When a correlation or regression coefficient is presented, is it accurate?** (Consider: Outliers, R^2, range restriction)			
11. **When a regression coefficient is presented, does the regression model fit the data?** (Consider: R^2)			
12. **When estimates are presented, are confidence intervals provided? If so, are they sufficiently small?**			

Big data, machine learning, and artificial neural networks



As discussed in the previous module, ***Big Data*** refers to extensive, diverse, and rapidly changing data sets. Walmart, for instance, processes more than a million transactions per hour (14). Such data are on a scale quite different from what most analysts are used to. For this reason, new data analytics techniques have emerged, using machine learning and artificial neural networks. Most of these techniques are based on the principle of artificial intelligence (AI). The e are many definitions of AI, but in general it means a program using data to build a model of some aspect of the world. This model is then used to make informed decisions and predictions about future events. (15) Machine learning is an application of AI that enables computer systems to automatically learn, adapt, and improve from experience (predict, test, and revise) and pattern recognition – without being explicitly programmed by a human being. The underlying principle of artificial neural networks is similar, but in this case, learning occurs from weblike connections among a large number of small processor units. These techniques mimic the way humans learn by refining the algorithms used to autonomously observe and analyze data

Most organizations, however, do not have big data – they don't need to analyse 100+ terabytes of heterogeneous data per day in real time, and thus "conventional" statistical techniques suffice. In addition, Big Data and AI technology raise serious social, ethical, and political concerns. For example, the output of Big Data and AI techniques is based on extremely complex mathematical algorithms, often protected by copyright, and thus act as a black box. Moreover, in some cases, even the developers themselves don't know exactly how their techniques work. As a result, algorithms can have hidden biases, inadvertently introducing gender or racial biases into the decision-making process. Big Data and AI can be very helpful, but there remains a danger of being blinded by the glamour of these new fields, failing to recognize their limitations. Critical appraisal and attention to data quality are as important when working with Big Data as when working with small data.

Finally: Organizational evidence – quality versus purpose

A last word on organizational evidence, the raw data that comprise it, and the issue of quality: the applications of organizational data and the people who apply them are very diverse. For example, data can be gathered by the government to assess regulatory compliance, by senior executives to make strategic decisions, by supervisors to make operational decisions, or they can be used to identify problems that require immediate attention (e.g., equipment breakdown, security breach). The quality of data in terms of reliability and validity will vary depending on their use and the people who use them.

Regulatory use, for example, is subject to the highest quality standards because such data are used to evaluate compliance with legal requirements. Data alerting us to a possible problem, however, require a different standard – such data first and foremost need to be timely, so we accept that they are biased toward false positives – finding a relationship between data when none really exists.

For example, a hypersensitive smoke detector – triggering sometimes a false alarm – is preferable to one that requires a huge fi e before sounding. In the middle of the quality spectrum are data used for strategic and tactical decisions. The people who use such data may differ on how valid and reliable they think the data should be. When critically appraising organizational data, we must keep in mind that they are used by different kinds of people for different kinds of decisions requiring different kinds of quality. Consulting additional evidence sources (e.g., subject matter experts, relevant stakeholders) helps you to determine what level of data quality you need and thus the extent to which you need to scrutinize the data. As explained in Module 1, evidence-based management is about critically appraised evidence from multiple sources. Thus, organizational data, big or small, should not be the sole source of evidence in any decision-making process.

We started this module by emphasizing that uncritical use of organizational evidence can lead to unnecessary organizational risks. As a management practitioner, you need to ask not only *What do the data say?* but also such critical follow-up questions as *Where do the data come from?, What kind of analyses were conducted?*, and *How confident should we be in the findings?*

We then discussed ten common barriers that reduce the trustworthiness of organizational evidence.

The first barrier is the ***absence of a logic model***, that is, a short narrative explaining why or when a problem occurs and how this leads to a particular outcome. Formulating a logic model prevents a fishing expedition through voluminous amounts of data. Instead, a logic model helps focus our assessments and analysis on meaningful relationships.

The second barrier is ***irrelevant data***. In general, organizations are better off with a small number of relevant and accurate metrics – avoiding a large number of fancy but irrelevant ones. Important indicators of relevance are whether your organizational data have a key place in your logic model and whether metrics and key performance indicators are actionable.

The third barrier concerns ***inaccurate data***. Organizational data may represent facts, but all kinds of biases and distortions can slip into them. An important question is, *How were the data collected, processed, and reported?* As a rule, because of human error, incentives to manipulate the data, or even a simple lack of understanding of the data, inaccuracy increases wherever people (not IT) collect, interpret, or summarize the data.

The fourth barrier concerns ***missing contextual information***. Without context (like who, what, and where or the base rate), organizational evidence can be misleading.

The fifth barrier is ***measurement error***. Even if something is measured perfectly, there always will be a measurement error because the indicators used introduce errors, for example, if the person taking measurements does not use the measurement tool properly. For this reason, you should always ask, *How is this measured? and Is that a reliable way to measure?*

The sixth barrier is the ***small number problem***: when we use a small sample, the metrics and KPIs based on that sample are particularly likely to deviate from the “true” value.

The seventh barrier is ***confusion involving percentages and averages***. A difference or change expressed in a percentage can be relative or absolute. In most cases, relative change – expressed as a percentage of the base value – is a more accurate metric than absolute change. We also explained that there are three ways to calculate an average: the mean, the median, and the mode. In some cases, they are identical, but often they are not. An average combines a large amount of data and represents it with a single number. As a result, averages tend to obscure the actual variation in the data. Thus, whenever a metric represents an average, you should look at the standard deviation, a measure that tells us how much the data deviate from the average.

The eighth barrier is ***misleading graphs***. Management information systems often include advanced dashboards and visualization tools that present data in appealing graphs and charts. Such visualizations can be misleading. Well-known examples are graphs that omit the baseline, manipulate the axes, or present cumulative rather than interval data.

The ninth barrier concerns *correlations and regressions and the problem of overfitting the data*. To get a good idea of how strongly two metrics are correlated, we can look at the r^2 (pronounced "r-squared"). Another aspect to consider when judging correlations is whether there is range restriction – which can lower the correlation actually observed. When regression coefficients are reported, it is recommended to check the residual plot, a graph that shows if the observed (true) values in a data set differ from the values as predicted by the regression (also referred to as *goodness of fit*). If a residual plot is not available, we can use the multiple R^2 instead.

The tenth barrier is *wide confidence intervals*. A confidence interval provides the upper and lower boundaries between which we expect the true value to fall. A 95% CI is considered too wide if the decision you would make based on the value of the lower boundary of the interval would be different from the decision you would make based on the value of the upper one.

At the end of this module, we discussed Big Data and advanced data analytics techniques such as artificial intelligence, machine learning, and artificial neural networks. Most organizations, however, do not have big data, so conventional statistical techniques suffice

Finally, when critically appraising organizational data, we must keep in mind that they are used by different kinds of people for different kinds of decisions requiring different kinds of quality. Consulting additional evidence sources helps you to determine what level of data quality you need and thus the extent to which you need to scrutinize the data.

Podcast: APPRAISE Evidence from the Organization

In this podcast host Karen Plum discusses with Martin Walker, Director of Banking & Finance at CEBMa, Jeroen Stouten, Professor of Organisational Psychology at KU Leuven University, Ravishanker Jonnalagadda, Senior Expert Data Science, People Analytics , Novartis Healthcare, and Eric Barends, Managing Director of CEBMa, evidence gathered from the organisation to aid our multi-source evidenced-based decision making.

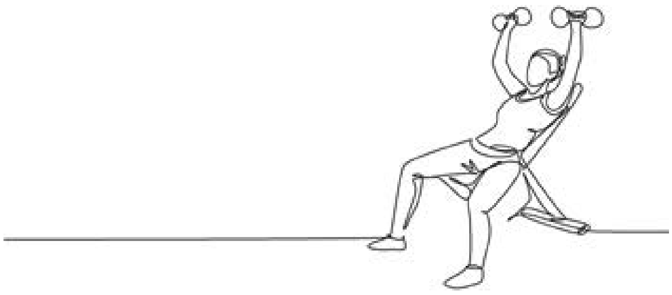
In this podcast we continue to consider some of the wider aspects associated with the use of organisational data. Why we can't take it's quality for granted; how it can be highly misleading if we don't know where it comes from and how it's created, and how people can continue using it for years in ignorance of its shortcomings.

The excitement around Big Data and data analytics is also discussed, to see whether these are all they are cracked up to be! And we look at the complexity associated with key performance indicators (KPIs) which again can lock us into repetitive behaviour, without questioning their value.

Once again, the importance of taking decisions based on multiple sources of evidence is reinforced.

https://evidencebasedmanagement.buzzsprout.com

Exercises Module



Exercise 8.1: Your company's metrics and key performance indicators: How relevant and reliable are they?

1. Make an overview of the metrics and KPI's available to you and managers in your organization.
2. Describe how these metrics/KPI's are made available (e.g., spreadsheet, dashboard)
3. Which four metrics/KPI's do you consider to be the most relevant. Why?
4. Describe how these four metrics/KPI's are collected, processed, and reported.
5. How would you judge the accuracy and reliability of these metrics/ KPI's? Why?
6. What other metrics/KPI's that are not available would you consider to be relevant? How would they help you or the managers in your organization make better decisions?

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Exercise 9.1: From data to information

It is generally believed that the absence rate for older employees (>55) is higher than for their younger counterparts (< 30) due to lower physical health due to aging. Do your company's data support or refute this belief?

1. Make an overview of the metrics and KPI's available to you and managers in your organization.
2. Describe how these metrics/KPI's are made available (e.g., spreadsheet, dashboard)
3. Which four metrics/KPI's do you consider to be the most relevant. Why?
4. Describe how these four metrics/KPI's are collected, processed, and reported.
5. How would you judge the accuracy and reliability of these metrics/ KPI's? Why?
6. What other metrics/KPI's that are not available would you consider to be relevant? How would they help you or the managers in your organization make better decisions?

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/ instructor.

Suggestions for further reading

- *Misunderstanding the Nature of Company Performance: The halo effect and other business delusions*, Rosenzweig, CMR, 2007
- *Data, data everywhere: A special report on managing information*, Kenneth Cukier, The Economist, Feb 2010
- *Evidence Based Management, Using Organizational Facts*, Lex Donaldson, Oxford Handbook of EBMgt, 2012
- *Making Advanced Analytics Work For You*, Barton & Court, HBR, 2012
- *Beyond the hype: Big data concepts, methods, and analytics*, Gandomi & Haider, IJIM, 2015
- *Big Data: Forget Volume and Variety, Focus On Velocity*, Brent Dykes, Forbes, June 2017
- *The era of blind faith in big data must end*, Cathy O'Neil, TED Talk, Sep 7, 2017
- *The human insights missing from big data*, Tricia Wang, TED Talk, Aug 2, 2017

CEBMa center for

Open Learning Initiative

References

1. McAfee, A., Brynjolfsson, E., & Davenport, T. H. (2012). “Big data: The management revolution.” Harvard Business Review. Volume 90. Number 10. 60-68.
2. Calude, C. S., & Longo, G. (2017). “The deluge of spurious correlations in big data.” Foundations of Science. Volume 22. Number 3. 595-612.
3. Vigen, T. *Spurious Correlations*. http://www.tylervigen.com/spurious-correlations.
4. Herrmann, D. S. (2007). *Complete guide to security and privacy metrics: Measuring regulatory compliance, operational resilience, and ROI*. Auerbach. Boca Raton, FL.
5. Merritt Hawkins and Associates (2012). 2012 Review of Physician Recruiting Incentives. https://www.amnhealthcare.com/uploadedFiles/MainSite/Content/Healthcare_Industry_Insights/Industry_Research/mha2012incentivesurveyPDF.pdf.
6. US Bureau of Labor Statistics (2012). National Compensation Survey.
7. Blastland, M., & Dilnot, A. W. (2008). The tiger that isn’t: Seeing through a world of numbers. Profile Book LTD. London.
8. Donaldson, L., & Rousseau, D. M. (Ed.) (2012). *Evidence-based management (EBMgt) using organizational facts: The Oxford handbook of evidence-based management.* Oxford University Press. New York. Example is adapted from Donaldson & Rousseau’s work.
9. Maanen, H. V. *Lesson 9: Understanding statistics*. World Federation of Science Journalists Online Course in Science Journalism. www.wfsj.org/course/en/.
10. As explained in Module 7, according to Cohen’s rule of thumb, a small effect is one that is visible only through careful examination. A medium effect 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.
11. Levitin, D. (2016). *A field guide to lies and statistics: A neuroscientist on how to make sense of a complex world.* Penguin. London.
12. Blodget, H. (2013). “These two charts show why Apple’s stock price is collapsing.” Business Insider. http://www.businessinsider.com/two-charts-show-why-apple-stock-dropped-2013-4.
13. Levitin, D. (2013). *Discovering statistics using IBM SPSS*. Sage. London. Edition 4.
14. Marr, B. (January 23, 2017). *Data at Walmart: Real-time Insights from their 40+ petabyte data cloud*. https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/#509055a56c.
15. Sample, I. (October 18, 2017). “’It’s able to create knowledge itself’: Google unveils AI that learns on its own.” The Guardian. https://www.theguardian.com/science/2017/oct/18/its-able-to-create-knowledge-itself-google-unveils-ai-learns-all-on-its-own.

Learn by doing & Did I get this? Answers and Feedback

Did I get this? 9.1

1: C

2: E

3: A

4: D

5: B

A logic model starts with a core assumption that explains why a problem occurs and is then followed by logical steps that explain how this leads to a certain organizational outcome. You could consider "*When newly hired sales agents fail to meet their sales targets over time, they are more likely to focus on immediate sales and disregard customer needs*" an important assumption, especially when there is no organizational evidence available (yet). However, we would argue that the core assumption is "*Disregarding customer needs and preferences could negatively affect sales performance.*"

Did I get this? 9.2

a: Yes

b: Yes

c: No

d: Yes

e: No

f: No

Learn by doing 9.1

Scenario 1

A logic model was NOT used > In this scenario, the data were collected and analyzed without using a logic model. Thus, the conclusion that millennials have a more lenient attitude toward sick leave could be a type I error: finding a elationship when none really exists.

Scenario 2

A logic model was used > In this example, the data were collected and analyzed on the basis of a logic model that looks like this:

1. Many managers in the company lack training in basic management skills.
2. Therefore, some managers lack essential people skills.
3. As a result, some employees don't feel appreciated and valued.
4. Thus, some employees leave the company.
5. Therefore, the company's turnover rate is higher than average.
6. This results in extra costs on selection and recruitment.

Did I get this? 9.3

A logic model was NOT used > In this scenario, the data were collected and analyzed without using a logic model. Thus, the relationship of gender, age, marital status, communication skills, sense of humor, car preference to bankers' perceptions of risk could be nonexistent or irrelevant.

Did I get this? 9.4

The goal of the sales campaign is to boost the number of sales of the new candy bar. The number of shares on Facebook is therefore not necessarily a relevant indicator. After all, watching and sharing a video is not the same as buying the candy bar. The number of candy bars sold would be a more relevant indicator.

Did I get this? 9.5

c,d: Correct > The norm set by the company's vice president should be based on a clear rule or (objective) standard. Otherwise, the score on the 3-point scale is prone to bias and may be inaccurate. The same is true for the summary and interpretation of scores by the channel marketing managers.

a,b,e: Incorrect

Learn by doing 9.2

1a: Correct

1b,c: Incorrect

When you spread out this amount over 5 years and divide it by the number of employees, it equals only $20 per employee. Spread this amount across 52 weeks a year for 5 years, and it equals 3 cents per week. Would you be able to find childca e for 38 cents a week?

2b: Correct

2a,c: Incorrect

After all, it depends on the total number of transactions. If the total number of transactions were, for example, 4.8 million per year, representing a total sum of $46.5 billion, a loss of $1.5 million would not be considered a big number.

Did I get this? 9.6

Our answer: We would like to know the banks' total number of clients. After all, a loss of 100 clients on a total of 1,000 is a bigger problem than a loss of 1,000 clients on a total of 6 million. In addition, we would like to know the profitability of the lost clients. A loss of 1,000 clients with small checking accounts with high overhead costs could even lead to a reduction of costs, whereas the loss of 100 billionaires with large checking accounts could create a serious risk to the bank's solvency.

Learn by doing 9.3

1: C

2: B

3: D

4: A

Learn by doing 9.4

There are many reasons why measurement error may occur. Some managers may be wearing a heavy suit while others are wearing jeans and a light shirt. Some might have a two-ounce candy bar in their pocket or a heavy set of keys. Some might just have a heavy lunch while others are on a diet. All these seemingly trivial details can have a significant impact on their measu ed weight.

Learn by doing 9.5

a, d, e: Correct

Learn by doing 9.6

Scenario 1

a,c: Incorrect

b: Correct

The larger hospital represents a larger sample size (45 babies) than the smaller hospital (15 babies). The law of large numbers states that a larger sample size will be nearer to the true value of the whole population (50% boys) than a smaller sample size. Thus, the smaller hospital will record *more* days on which more than 60 percent of the babies born were boys.

Scenario 2

a,b: Incorrect

c: Correct

It is hard to determine which business unit has a better retention. Small business units represent smaller samples, so because of the small number problem, the 20 percent turnover may be substantially lower (or higher).

Did I get this? 9.7

a,b: Incorrect

c: Correct

It is hard to determine whether the absenteeism rate at the 65 stores is too high, this depends on the number of employees that they employ. If the number of employees is much smaller than in the other stores, the rate of 6% may be distorted due to the small number problem and thus be substantial lower (or higher).

Learn by doing 9.7

a,b: Incorrect

c: Correct

Most people will be inclined to say no, because a 200 percent mortality rate increase sounds pretty dramatic. However, you need more information because this depends on the base value. If the mortality rate of the existing drugs is only 1 in 350,000 patients (0.000003 percent), a relative increase of 200 percent means an absolute increase of only 2 in 350,000 patients (0.000006 percent). In all, the new drug sounds like it has better outcomes, especially as a patient's improvement in health would be substantial.

Did I get this? 9.8

a,b,c: Incorrect

d: Correct

In 2017, the number of medication errors in unit A was 1,050 out of 50,000 patients, which is 2 percent. We don't know, however, how many patients were treated and how many medication errors were made in unit B. For example, if unit B treated 12,000 patients of whom 846 received the wrong medication, then an increase of 15 percent results in a total of 966 medication errors in 2017. However, in that case, 966 errors out of 12,000 patients is 8 percent, which would be much higher than the 2 percent of unit A.

Learn by doing 9.8

a,b,c: Incorrect

d: Correct

The mean, the median, and the mode have the same value: 40.

Did I get this? 9.9

Scenario 1

a: Correct

b,c,d: Incorrect

The mean is 60, whereas the median is 70 and the mode is 72. If you look at the range of the numbers, which vary from 31 to 72, you see that the mean is indeed a better representation than the median or the mode.

Scenario 2

a,b,d: Incorrect

c: Correct

The mean is 60, whereas the median and the mode are both 44. If you look at the range of the numbers, you see that the median and the mode are indeed a better representation than the mean.

Learn by doing 9.9

a,b,d: Incorrect

c: Correct

The rule of thumb states that about 95 percent of the data will be covered by two standard deviations plus or minus the average. Thus, in this example, a standard deviation of 5 tells us that 95 percent of the employees are between 25 years (35 – [2 × 5]) and 45 years old (35 + [2 × 5]).

Did I get this? 9.10

a,b: Correct

c,d: Incorrect

The metric of 10 percent is not a good representation of the company's true absenteeism rate, because the standard deviation of 5 indicates that 95 percent of the company's true absenteeism rate is somewhere between 0 and 20 percent. This means it is unclear whether the company's absenteeism rate is on par with the average in the sector.

Learn by doing 9.10

a: Correct

b,c,d: Incorrect

A decrease of 5.5 medical errors with a standard deviation of 20 equals 5.5/20 = 0.26, so this is a small decrease.

Did I get this? 9.11

Graph 1

a: Correct

b,c,d: Incorrect

By omitting the baseline, the graph suggests that managers are much more in favor of the merger than are employees and stakeholders.

Graph 2

a: Correct

b,c,d: Incorrect

This is not a misleading graph.

Did I get this? 9.12

a: Incorrect. You are correct that the correlation is statistically significant, but even a statistically significant cor elation may be practically irrelevant.

b: Not quite right. You are correct that the correlation may be spurious, especially when a voluminous amount of data is captured and exhaustively analyzed. However, that is not the main reason you should not support the sales director’s recommendation.

c: Correct. The r-squared of the correlation is 0.014 (0.12 × 0.12), indicating that only 1.4 percent of the variation in the likelihood that a client will file a claim (inc ease or decrease) can be explained by the difference in clients’ weight.

d: Incorrect (see C)

Did I get this? 9.13

a: Correct. The blue dots in the scatter plot are closely distributed around the red line, suggesting a strong correlation. However, there are also seven dots that lie far outside the line. Because outliers can easily deflate the cor elation coefficient, these seven outliers may have a fected the correlation as calculated by the data analyst.

b: Incorrect. See A

c: Incorrect. There is always a chance that a correlation found is spurious, but given that your analysis of the correlation between education and claims was not a hunch but was based on previous empirical studies, this is not likely.

d: Incorrect, see A

Learn by doing 9.11

This is because their lawyers’ IQ is probably on the high side, perhaps between 125 and 140, whereas the IQ of the general population ranges from 70 to 130. As a result, the outcome is distorted due to range restriction.

Did I get this? 9.14

a: Correct

b,c,d: Incorrect

Did I get this? 9.15

a: Not quite right > The multiple R2 is the proportion of the variation in the outcome variable (increase or decrease of the number of claims) that is predicted by the predictor variable (level of education). However, that does not mean that you should multiply this proportion with the regression coefficient

b: Incorrect > A multiple R^2 of 0.5 equates to 50 percent, not 5 percent.

c: Correct

Learn by doing 9.12

a,c,d: Incorrect

b: Correct

The confidence interval indicates that there is a 95 percent confidence that the true number of defects falls somewhere between 5.3 and 6.1 per 1,000 cars, which is sufficiently p ecise.

Did I get this? 9.16

a: Correct

b,c; Incorrect

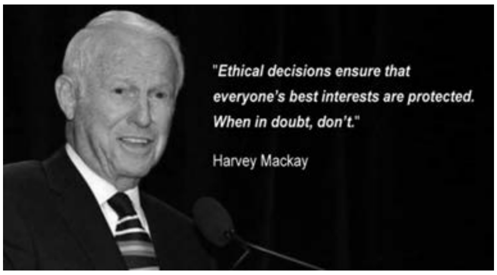
Of course, there are still many questions to answer, such as what type of absence and where. However, the confidence interval indicates that the e is a 95 percent confidence that the true absenteeism rate falls somewhere between 3.2 and 7.2, which is too wide: it indicates that the true absenteeism rate may even have decreased. In that case, there is no problem that needs solving.

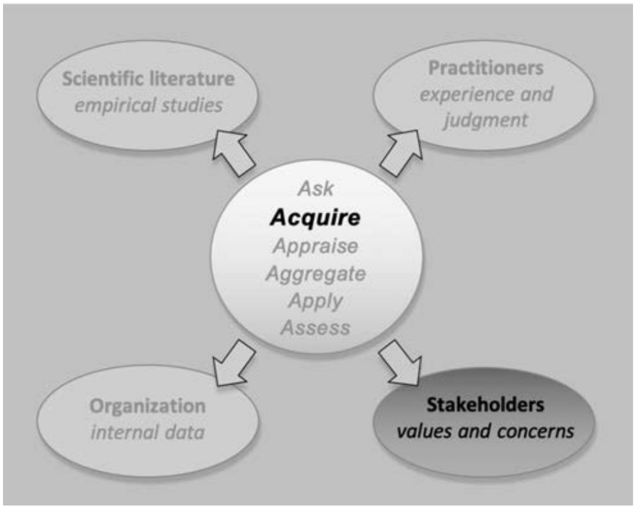
MODULE 10 | ACQUIRE Evidence from Stakeholders

Learning objective:

- Acquire evidence from relevant stakeholders in a valid and reliable way.

Introduction





Consider these two real-life examples:

In October 2016, KLM – a Dutch airline company – decided to assign one less flight attendant in economy class on 40% of its long-haul flights. According to KLM's management, the change would produce a productivity gain of 4% while having minimum impact on customers. KLM's flight attendants, however, strongly felt that reducing the number of crew members would increase their workload considerably, creating problems with customer service. KLM's crew members announced a 24-hour strike, which would cost the company more than $10 million. After flight attendants and the airline reached an agreement, the strike was called off. KLM's management granted flight attendants a 3.5% salary increase, and cancelled the reduction in staff on long-haul flights. (1)

In September 2017, Uber – a global taxi company from the US – was stripped of its London licence, dealing a serious blow to one of Silicon Valley's fastest rising companies. Uber's licence was rejected on the basis that the firm failed to show corporate responsibility. Especially in London, Uber faced serious criticism from unions, lawmakers and cab drivers over its working conditions. London mayor, Sadiq Khan, backed the decision, saying "… all companies in London must play by the rules and adhere to the high standards we expect." Uber's chief executive, Dara Khosrowshahi, said that he disagreed with the decision because it was based on past behavior. (2)

We define stakeholders as people (individuals or groups) whose interests affect or are affected by an organization's decision and its outcomes. Stakeholders can be inside the organization as in the case of employees facing downsizing and restructuring or IT staff responsible for implementing a new on-line purchasing system. Stakeholders also can come from outside the organization as in the case of equipment suppliers subject to a firm s new requirements on energy use, or the neighbors of a manufacturing plant that cuts back on its frequency of garbage pickup. Given their potential connection to a decision's consequences, stakeholder interests, values, and concerns are essential considerations in organizational decisions, making them an important source of evidence.

In the context of organizational or managerial decisions, evidence from stakeholders pertains to both practical and ethical considerations. Practical considerations arise from the effect, appropriateness, or fairness of a decision *in the eyes of its stakeholders*. The impact a decision has on stakeholders, whether objective or perceived, can affect their willingness to accept (or support) that decision. In the example of KLM, the cabin crew – an important and powerful group of stakeholders – perceived the company's decision to reduce the number of flight attendants as considerably increasing their workload, and as harmful to passengers. As a result, the company's decision wasn't implemented. Evidence from stakeholders, however, can also address ethical considerations, particularly in terms of the distribution of harms a decision generates relative to its potential benefits. In the example of Uber, many people perceived the company's business strategy and market practices as disproportionality harmful to competitors. As a result, policymakers and regulators – powerful external stakeholders – regarded the company's practices as unethical, demonstrating a lack of corporate responsibility.

In this module, we explain how to identify a company's most relevant stakeholders. We also discuss methods for exploring stakeholder interests and concerns, and describe how paying attention to both practical and ethical aspects of the decision process can improve the quality of your decisions. Before we start, however, we must dedicate a few words to the difference between stakeholder evidence and other sources of evidence.

Not all evidence from stakeholders is stakeholder evidence

Not all evidence from stakeholders is stakeholder evidence. Sometimes evidence from stakeholders actually represents organizational data or practitioner judgment. Consider for example the case where the distribution center of a large food retailer experiences high error rates. Senior management suggests that introducing Lean Management practices would optimize the center's operational processes and thus reduce errors. Consultation with the center's staff, however, suggests this change wouldn't have the intended effect: Staff perceive the problem as the result of recently hired low skilled employees rather than inefficient processes. They point to the increase in error rates over the past year, coinciding with a cost cutting program prohibiting the hiring of more skilled (i.e. more expensive) employees.

In this example, the various kinds of information staff might provide represent different kinds of evidence. First, the staff can provide stakeholder evidence (i.e. how their interests might be affected by senior management's decision), which might be that their preferred way of working may change if Lean Management is introduced. Second, staff can provide organizational evidence in the form of information regarding objectifiable facts elated to the assumed problem (high error rates due to low skilled employees) and the likely effectiveness of the proposed solution (Lean Management practices).

Now consider the evidence from the stakeholders in the KLM example. In this case, the stakeholders ***strongly felt*** that the decision would have negative effect on their *perceived* workload. The same counts for the policymakers and regulators in the Uber example: they perceived Uber's practices as conflicting with the city's ethics and values. In all these cases, the evidence from stakeholders concerns *subjective feelings* and *perceptions* that cannot be considered as objectifiable facts regarding an assumed problem or proposed solution. Such information is, however, relevant to the decision and constitutes stakeholder evidence.

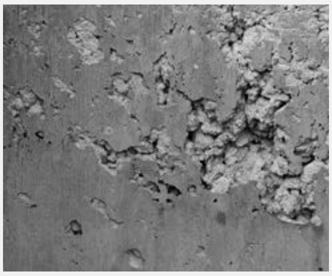
Learn by doing 10.1

Consider the following two scenarios

Scenario 1

The dean of a faculty of a British university notices that the number of research publications has dropped with 20 percent over the past 5 years. Because the number of publications is the most important evaluation criteria of the university's ranking in the Times Higher Education World University Ranking, he decides to introduce a financial incentive: for every publication in a top journal a faculty member will be awarded a bonus of £200. When he discusses his decision with the faculty members, they state that they don't feel recognized by the faculty's management: when their research is published in a top journal, they seldom receive a compliment, a private note or email, or recognition of their achievement in the faculty's newsletters. The dean, however, dismisses their complaints and decides to introduce the financial incentive anyway.

Scenario 2



In the building and construction industry, rough pitted surfaces or voids in concrete are referred to as 'honeycombs'. Honeycombs are resulting from incomplete filling of the concrete against the framing, often caused by using concrete that is too stiff or by not vibrating it sufficiently after it has been poured. A Canadian building and construction company notices that the number of honeycombs in the concrete has substantially increased over the past year. The manager thinks this is due to the increased number of young and inexperienced workers the company has hired in the past year. He therefore decides to give these workers extra training. When he discusses this decision with some of the workers, however, they point out that the increase of honeycombs coincides with the introduction of a new pouring technique. Based on this insight, the manager decides to first ask managers from companies that use the same technique whether they experience similar problems

Which scenario involves an example of stakeholder evidence?

a) Only scenario 1 concerns stakeholder evidence
b) Only scenario 2 concerns stakeholder evidence
c) Both scenarios concern stakeholder evidence
d) None of the scenarios concern stakeholder evidence

Did I get this 10.1

Scenario 1

The dean of the faculty of a British university that we described above decides to take the feelings of the faculty members seriously. He reverses his decision to introduces financial incentives, and instead publicly recognizes faculty members when their research is published in a top journal, gives compliments, sends them a private note, and dedicates a short article about each new publication in the university's newsletter. In the next 6 months, the number of research publications increases. To further stimulate the number of publications, he decides to reintroduce his idea of financial incentives. When he discusses his decision with the faculty members, they advice him to increase the bonus from £200 to £1000, because a recent published meta-analysis has demonstrated that a financial incentive should be substantial rather than symbolic.

Scenario 2

The manager of the Canadian building and construction company that we described above learns from his colleagues from companies that use the same pouring technique that they did not experience an increase of honeycombs. He therefore decides to introduce mandatory training for all workers. This training will be provided outside regular working hours, in the workers' own time. The workers, however strongly oppose this decision: they feel manager's decision is not fair and will increase their workload. Some workers even call for a strike.

Which scenario involves an example of stakeholder evidence?

a) Only scenario 1 concerns stakeholder evidence
b) Only scenario 2 concerns stakeholder evidence
c) Both scenarios concern stakeholder evidence
d) None of the scenarios concern stakeholder evidence

Why is this distinction important?

With practitioner-, organizational-, and scientific evidence, the key issue in critical appraisal is whether we can trust the evidence. As explained throughout the book, we should always critically appraise the evidence. We ask questions about how the organizational data were collected or the outcome measured, whether practitioners consulted had sufficient experience and the quality of the feedback they received, if the scientific evidence came from a study using a control group and had a large enough sample, etc. In contrast, stakeholder evidence involves subjective feelings and perceptions – so reliability and validity is not the major issue. Critical appraisal instead focuses on whether stakeholder evidence accurately represents the feelings and perceptions of all stakeholders in that group. A key idea is that the type of evidence determines how the evidence is critically appraised. As we will see in the next module, when we are dealing with the critical appraisal of stakeholder evidence, representativeness is a major indicator of its quality.

Types of stakeholders

Stakeholders take a variety of forms. As a result, there are many ways of classifying stakeholders. The most common distinction is probably between internal stakeholders (e.g. employees, managers and board members), connected stakeholders (e.g. customers, suppliers, distributors, financiers, and shareholders), and external stakeholders (e.g. regulators, government, professional bodies, local communities, and the society at large).

Another relevant distinction is that between direct vs indirect stakeholders, reflecting whether a decision has a direct impact on the stakeholder's interests or an indirect impact through the actions of other stakeholder groups. For example, a call center agent working for a retail company that decides to open up all stores one hour later on workdays is not directly affected by this decision, but he/she may be confronted by customers who are. This makes the agent an indirect internal stakeholder, and the customer a direct external stakeholder. Another distinction is that between primary and secondary stakeholders, which is based on the company's responsibility towards the stakeholder. For example, current employees, customers, and suppliers are primary stakeholders, whereas future employees, regulators, and the local community are secondary stakeholders.

Learn by doing 10.2

Consider the following scenario

The dean of an American community college for vocational education (a school that prepares 16 to 18 year olds to become an electrician, hair dresser, mechanic, etc.) noticed that during information sessions, parents often ask questions about student safety (e.g. bullying, harassment, intimidation, abuse, verbal and physical aggression, alcohol, drugs, etc.). She usually states that the number of safety related incidents is very low, but this answer is based only on personal experience rather than trustworthy evidence. All teachers and students are required to report all safety related incidents, so she therefore consults the school's management information system. The system indicates that in the past semester there were only 12 safety related incidents.

Although this number is rather low, she decides to place walk-through metal detectors for checking students for weapons and hire security guards that are allowed to randomly stop and search students for possession of drugs.

In this scenario, what type of stakeholders are the following groups?

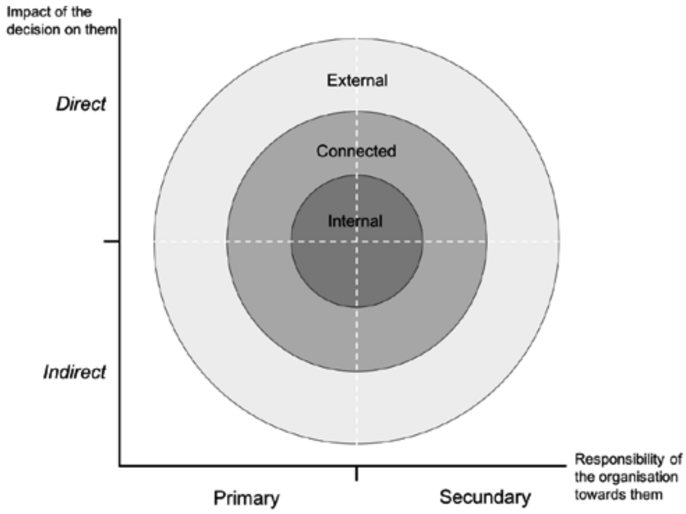
a) Students and staff

b) Students' parents

c) Organizations (e.g. highs schools) that refer students to the school

Mapping stakeholders related to a decision or problem

A stakeholder map is a useful tool to overcome the decision maker's biased consideration of a decision's implications for other people. This map illustrates the potential array of stakeholders related to a specific decision. Because decision makers often differ in their views of whom these stakeholders might be, it helps to make stakeholder mapping a public process and seek advice from experienced practitioners in doing so. Below is a map to identify stakeholder groups affected by a decision. Note that a stakeholder's position on the map (e.g. direct vs indirect) depends on the decision being made. When a decision changes, so may the position of the stakeholder(s). (3)



Organizational decisions often have lots of stakeholders, both inside and outside the organization. In some situations, this may lead to 'information avoidance' – the tendency for decision makers to avoid paying attention to stakeholder groups for whom less information is available. In addition, some stakeholders may be 'out of sight' and not come to mind initially, particularly when a decision has indirect or long-term effects. Awareness and concern for possible effects on stakeholders (or lack thereof) is a focus of much debate regarding corporate social responsibility. (4)

Did I get this 10.2

Consider the following scenario

The board of an American chemical company that produces insecticide for use in commercial applications has developed a new waste management program. Until now, all the company's waste that results from the chemical manufacturing process is shipped to China and stored at a landfill in the province of Jiangsu. Because China will step up efforts to end the storage of hazardous chemical waste over the next six years, the new waste management program proposes alternative *storage options, one of which is building a landfill site close to the company's main processing location. This landfill will be a so called "sanitary landfill" where the company's waste will be separated from the surrounding environment using a system of layers designed to allow waste to decompose safely. In addition, the landfill will be closely monitored for at least 50 years to ensure the surrounding environment is safe from contamination. The local community of a nearby town, however, fear that the storage may have harmful effects on the environment and peoples' health. The company's shareholders caution the board to take a careful and conscientious decision, as this could potentially harm the company's reputation and market value. Some of the company's senior managers are also skeptical about the idea of local waste storage, in particular the coordinating manager that is responsible for the disposal of the company's waste. However, rumor has it that this manager is biased, because she used to work for the shipping company that ships the waste to China. In addition, if the company's waste is stored locally, the manager may lose her job.*

Based on this scenario, place the groups of stakeholders in the correct quadrant in the table.

	Internal	External
Direct		
Indirect		

a) ***The coordinating manager***

b) ***The local community***

c) ***The shipping company***

d) ***The company's shareholders***

Who are the most relevant stakeholders?

After mapping all stakeholders for a decision, the next step is to determine the most relevant. A stakeholder's relevance is determined by two variables: the extent to which the stakeholder's interests are affected by the decision (harms and benefits), and the extent to which the stakeholder can affect the decision (power to influence).

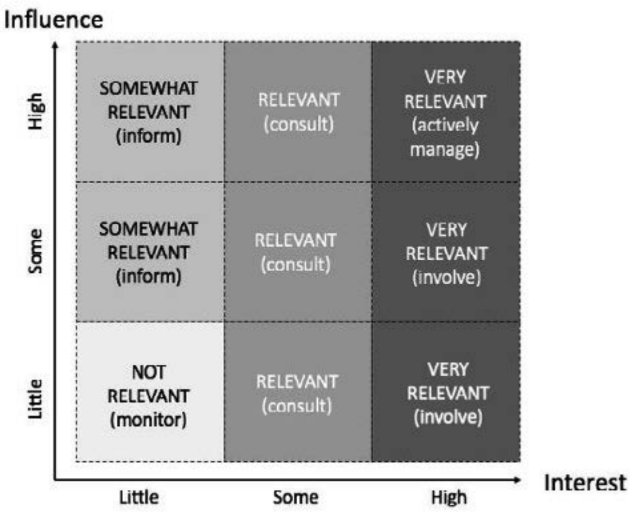
Harms and benefits

Harms and benefits can di fer considerably across stakeholders. For example, some managers, employees and shareholders may benefit f om a decision to close a line of business while certain customers and suppliers may be harmed. The key idea is that some stakeholder groups may experience gains and benefits f om an organizational decision while others bear more costs and harms.

In general, the more a stakeholder group stands to benefit or lose by a decision, the stronger their interest in the decision tends to be. When identifying stakeholders' interests, keep in mind that the focus is on their perception of harms and benefits, not whether these harms and benefits will actually occur.

Finally, the more costs and harms stakeholders perceive, the more likely they may seek to influence the decision-making process. Nonetheless, stakeholders whose interests are affected by a decision may be unaware of it or otherwise unable to exert influence over it. Some stakeholders may only surface relative to a decision after it has been made. Nonetheless, the ultimate quality of the decision may be affected by how all stakeholders are affected by it.

Power to influence



Whether stakeholders can exert influence over a decision depends on the power they can exercise. This power can derive either from the stakeholders' authority and position within the organization or their ability to entice/convince/persuade others with power to influence the decision-making process. In either case, the more power stakeholders have, the better they will be able to positively or negatively affect the decision.

In general, stakeholders with a high level of interest and power can be considered particularly relevant stakeholders. As illustrated in the diagram left, you can assess a stakeholder's relevance by mapping their level of power and interest.

Learn by doing 10.3

Consider the following scenario we used in the previous exercise

The board of an American chemical company considers building a landfill site where the company's waste will be stored. Although all necessary measures will be taken to ensure the surrounding environment is safe from contamination, several stakeholders have some serious concerns:

- ***The local community:*** fears that the storage may have harmful effects on the environment and peoples' health.
- ***The company's shareholders:*** think the decision could potentially harm the company's reputation and market value.
- ***The coordinating manager:*** fears that if the company's waste is stored locally, she may lose her job.
- ***The shipping company:*** may lose their contract to ship the company's waste to China

Because of these concerns, the board decides to make an overview of all stakeholders. In addition to the stakeholders mentioned above, they also identify the following stakeholders:

- ***The legal department:*** there are a huge number of legal requirements for the disposal of chemical waste, so this may create extra work them
- ***The Waste Disposal Authority:*** is very strict when it comes to law abidance and is known for its harsh economic sanctions
- ***The local press:*** is the mouthpiece of the local community, can create negative media attention that may harm the company's reputation
- ***Resellers:*** may choose another supplier when the company's reputation is damaged
- ***Employees:*** most of them are from the local community, so some people may hold the company's decision against them

Based on this scenario, place the groups of stakeholders to the correct quadrants in the table

INFLUENCE	High	Some	Little
INTEREST			
High			
Some			
Little			

a) ***The shipping company***
b) ***The regulating authority***
c) ***Resellers***
d) ***The legal department***
e) ***The coordinating manager***
f) ***The local press***
g) ***Employees***
h) ***The shareholders***
i) ***The local community***

Did I get this 10.3

Based on the outcome of the previous exercise, how would the company's board rank order the relevance of the stakeholders (most relevant on top)?

A. ***The shipping company***
B. ***The regulating authority***
C. ***The resellers***
D. ***The legal department***
E. ***The coordinating manager***
F. ***The local press***
G. ***The employees***
H. ***The shareholders***
I. ***The local community***

Ethical considerations

As we explained in the introduction of this module, evidence from stakeholders can address both practical and ethical considerations in organizational or managerial decisions. Practical considerations arise from stakeholders' level of power and interest, indicative of the effect they may have on the decision process itself. Ethical considerations arise from the distribution of a decision's potential harms relative to its benefits. In the diagram above, stakeholders with high interests but little influence a e often considered by decision makers to be less relevant to those with high influence. F om an ethical perspective, however, this is problematic.

Learn by doing 10.4

In the 'Did I Get This' exercise above, we ranked-order the relevance of the stakeholders. From an ethical perspective, however, you may not agree. How would you, taking ethical considerations in mind, rank order the relevance of the stakeholders?

Example

Between 2005 and 2007, Goldman Sachs, one of the most prestigious financial firms in the world, sold its clients so called mortgage-backed securities (MBSes), a type of investment that is secured by a mortgage – allowing its clients to benefit from the booming housing market, without having to buy an actual house. In 2006 however, home prices began to slide and an increasing number of homeowners struggled to pay their monthly mortgage. As a result, the value of MBSes declined, and the bank's clients were likely to lose their investment. However, despite these worrying signs, Goldman Sachs did not alert the 200,000 clients who had bought MBSes, instead they secretly made a bet that the clients would lose their money. (5) *As a result, the bank made a profit of billions of dollars. Although the bank stated that they were simply following normal business practices and did nothing illegal, their decision to put corporate profit before stakeholder interest was considered by many as unethical. As a result, Goldman Sachs had to pay up to $5 billion to settle claims of wrongdoing. Associate Attorney General Delery said that the settlement "makes clear that no institution may inflict this type of harm on the American public without serious consequences."* (6)

As a rule, regardless of whether they have the power to influence the decision-making process, when stakeholders' interests are affected by a decision, we consider them relevant.

In Module *Apply* we will discuss how you can account for ethical aspects when incorporating stakeholder evidence into the decision-making process.

Whom to ask?

We have explained how to identify stakeholder groups pertaining to a decision and how to determine the most relevant stakeholders – using a stakeholder map and a power/interest diagram. However, to fill out the map and diagram, we need to answer these four questions and sub-questions first

1. Who could affect this decision, its implementation, or its outcome?

- *What are their interests and concerns?*
- *What are their feelings and perceptions of this decision?*
- *How much influence do they have?*
- *Do they have legal rights to participate in this decision?*
- *How could they affect the decision?*
- *Do they have the power to block the decision or impede its implementation?*

2. Who could be affected by this decision?

- *What are their interests and concerns?*
- *What are their feelings and perceptions of this decision?*

3. Who may experience harm from this decision?

- *What are their interests and concerns?*
- *What are their feelings and perceptions of this decision?*
- *How well are their interests balanced with those who may benefit?*

4. Who may stand to benefit from this decision?

- *What are their interests and concerns?*
- *What are their feelings and perceptions of this decision?*
- *How well are their interests balanced with those who may be harmed?*

Using the answers to these questions, you should be able to fill out the stakeholder map and the power/interest diagram, and determine who are – given the decision at hand – the most relevant stakeholders. In general, these are the stakeholders to consult in the decision process.

Learn by doing 10.5

Read the scenarios and select the correct stakeholder(s)

Costco Wholesale is one of the largest American retail companies. In 2019, Costco's quarterly revenues were $34.74 billion and their stock was trading at an impressive $269.14 a share. In 2019 the company decided to raise its base wage from $13 an hour to $15 an hour. In contrast, the average retail worker in America makes between $7 and $9 an hour.

The reason for this decision is that the company's success comes from the high level of customer service offered by satisfied employees – by paying higher wages, the company attracts high quality workers. In addition, the company has low turnover and less labour conflicts because its employees are happier.1

Which three stakeholder groups have the power to directly influence this decision and/or its implementation?

a) The employees
b) The customers
c) The shareholders
d) The HR department
e) The legal department
f) The suppliers
g) The managers

Learn by doing 10.6

Read the scenarios and select the correct stakeholder(s)

Chick-fil-A, an American fast-food chain, has decided to pay up to $25,000 in tuition assistance to employees. As a result, in the past years, the company has paid more than $75 million in tuition to 53,000 employees. Interestingly, 90% of employees who received tuition assistance say they will keep working for the company. Moreover, 60% of the tuition recipients claim they could not have attended college without company help. Plus, 20% of Chick-fil-A scholarship recipients say they are the first in their families to attend college.

Apparently, treating employees well is paying off for Chick-fil-A, the company's sales grew by $1.1 billion in 2017; twice the growth rate of Wendy's and Burger King combined.[1]

Which three stakeholder groups may benefit the most from this decision?

a) The employees
b) The customers
c) The shareholders
d) The HR department
e) The society at large
f) The suppliers
g) The managers

Did I get this 10.4

Best Buy, a large American electronics retailer, is committed to reducing its environmental impact by reducing waste. For example, the company's customer service team members drive hybrid cars, which has saved 140,000 gallons of gas, the carbon equivalent of taking 263 cars off of the road for a year. In addition, Best Buy encourages its customers not to throw electronics away and claims to have collected two billion pounds of unwanted electronics and appliances for recycling. Finally, the company has set up 'Teen Tech Centers' that teach disadvantaged young Americans basic technology skills. This way, the Centers help reduce unemployment in America which has a serious shortage of vocational education.

Apparently, ethics are paying off for Best Buy. Notably, the company is now the only national electronics chain left in the United States. Its most famous competitor Radio Shack went out of business and other competitors like Circuit City are also long gone. (7)

Which stakeholder group may experience the most harm from these decisions?

a) The employees
b) The customers
c) The society at large
d) The shareholders
e) The competitors
f) The suppliers
g) The managers

However, sometimes it is not feasible (or desirable) to obtain evidence from all relevant stakeholders, for example when the number is simply too large or some stakeholders are inaccessible (e.g. geographic distance). In those cases, one could rely on 'key' stakeholders – opinion leaders who represent the interests of a stakeholder group – or experts knowledgeable regarding stakeholders' interests, values, and concerns.

For reasons of representativeness and legitimacy, however, we strongly recommend you to acquire evidence from a larger representative sample rather than a small group of individuals.

In addition, you should consider consulting secondary data, like scientific evidence regarding employee or customer perceptions, as other evidence sources may enhance your understanding of the perception and feelings of the employees and customers in your own organization.

Feelings and perceptions: how to ask?

You can acquire evidence from stakeholders in many ways. Numerous books and websites are informative about how to gather evidence in a valid and reliable way. They cover important aspects such as sampling procedures, methodology, and questionnaire development. In addition, the tools and techniques of marketing research are also useful in gathering evidence from stakeholders, particularly in the case of external stakeholders such as clients, consumer groups, or the general public. Negative sentiments spread like wildfire through social media, so for some companies it may be useful to monitor platforms such as Facebook, Twitter, and LinkedIn, and to follow key stakeholders active on social media. For most organizations, however, traditional quantitative and qualitative methods will suffice

As we discussed in previous modules, gathering evidence is not a fishing expedition – it starts with an assumed problem and/or a proposed solution. Before consulting stakeholders, it is important that you first clearly describe the problem to be solved. A good problem definition entails at least three elements:

1. The problem (or opportunity) itself, stated clearly and concisely: What? Who? Where? When?
2. Its (potential) organizational consequences
3. Its assumed major cause(s)

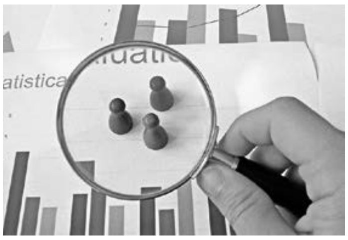
The same counts for the proposed solution – having a clear description of that solution is a prerequisite. Again, a good description entails at least three elements:

1. The solution itself, stated clearly and concisely: What? Who? Where? When?
2. Its (potential) effect on the problem and underlying cause(s)
3. Its costs and benefits

Quantitative methods

Quantitative methods are used to gather data (evidence) that can be quantified, that is, measured and written down with numbers. We usually obtain these data from surveys, structured interviews, tests, or voting systems. Quantitative methods are widely used to acquire evidence from practitioners. In Module 3 an overview of the most common methods is provided. When acquiring evidence from stakeholders, you can use quantitative methods to quantify their perceptions and feelings (how many stakeholders feel/perceive …). However, when it comes to exploring or identifying these feelings and perceptions (how do stakeholders feel/ perceive ...), qualitative methods are often more appropriate.

Qualitative Methods



Qualitative methods are used to gather data that cannot be scored or written down with numbers. We typically obtain these data from interviews, focus groups, or text analysis. Qualitative methods are often exploratory in nature: they are used to gain a better understanding of underlying feelings, opinions, or motivations. Thus, qualitative methods are particularly useful in obtaining stakeholder perspectives in their own words.

Focus groups

One of the most widely used methods to gather evidence from stakeholders are focus groups. A focus group is a set of six to ten people who are asked about their perceptions, feelings, opinions, or attitudes towards a product, service, idea, or – in case of an evidence-based approach – asked about an assumed problem or proposed solution.

The questions are asked by a skilled moderator in an interactive group setting where participants exchange point of views with other group members. The group needs to be large enough to generate rich discussion but not so large that some participants are left out.

Focus groups are a qualitative methodology where the researcher/moderator takes notes or makes recordings of the most important points obtained from the group. Decision makers can review the resultant transcripts or summary.

Focus groups typically yield data and insights less accessible without interaction between the group members. Members of the focus group should be selected carefully in order to obtain representative responses.

In addition, measures should be taken to prevent groupthink and other biases that may distort the outcome, as discussed in Module 3 and 4. A related method is a so called ‘public review process’ where stakeholder perspectives are obtained, discussed and vetted in a public session.

You can download a guideline on how to conduct focus groups here

Qualitative interviews

Interviews often are used to gather information from an individual stakeholder or group. Interviews can be structured, semi-structured, or unstructured. Structured interviews use a fixed format in which all questions are prepared beforehand and are asked in the same order. To ensure that answers can be reliably aggregated and comparisons made, all stakeholders are asked the same questions.

Most qualitative interviews, however, are unstructured or semi-structured. Unstructured interviews don’t use a predetermined questionnaire and may simply start with an opening question such as ‘Can you tell me how you feel about X?’ and then advance depending on the stakeholder’s response. (8)

Unstructured interviews are often difficult to conduct, as the lack of predetermined questions provides the interviewer little guidance on what to ask. For this reason, semi-structured interviews are perhaps the most widely used. This type of interview consists of a limited number of key questions that define the topic or issue to be explored while allowing the interviewer to explore relevant information not thought of beforehand.

Many books and guidelines describe how to conduct a qualitative interview. Generally, we recommend to use open-ended (i.e. questions that require more than a yes/no answer), neutral (non-value-laden or leading), and understandable questions. Formulating questions, however, is a process that requires particular attention. The choice of words to use in a question is critical – even small wording differences can substantially affect the answers people give. Module 3 provides tips to reduce comprehension error as a result of the question's wording.

When interviewing stakeholders, it is important to inform them in advance about why you need their input. You also need to set clear expectations regarding anonymity, confidentiality, and how stakeholder information will be used, as this increases the likelihood of honesty. (9)

Finally, we recommend you to record all interviews and focus group sessions, as this prevents against bias and provides a verifiable record of what was (and what was not) said. (10)

We started by pointing out that not all evidence from stakeholders is stakeholder evidence. Sometimes evidence from stakeholders actually represents organizational data or practitioner judgment. Stakeholder evidence concerns subjective feelings and perceptions that cannot be considered as objectifiable facts

Next, we discussed type of stakeholders. There are many ways of classifying stakeholders. The most common distinction is between internal stakeholders and external stakeholders. Another relevant distinction is that between direct vs indirect stakeholders, reflecting whether a decision has a direct impact or an indirect impact on the stakeholder's interests. Finally, we discussed the difference between primary and secondary stakeholders, which is based on the company's responsibility towards the stakeholder.

Organizational decisions often have lots of stakeholders, it is therefore recommended to draw up a 'stakeholder map' that illustrates the potential array of stakeholders related to a specific decision. Another useful tool is a power/interest diagram that illustrates 1) the extent to which the stakeholder's interests are affected by the decision (harms and benefits), and 2) the extent to which the stakeholder can affect the decision (power to influence)

In general, stakeholders with a high level of interest and power can be considered relevant stakeholders. Stakeholders with high interests but little influence a e often considered by decision makers to be less relevant. From an ethical perspective, however, this is problematic. As a rule, regardless of whether they have the power to influence the decision-making process and/or its implementation, when stakeholders' interests are affected by a decision, we consider them relevant.

Next, we discussed how you can acquire evidence from stakeholders. Gathering evidence, however, is not a fishing expedition – it starts with an assumed problem and/or a proposed solution. Before consulting stakeholders, it is therefore important that you first clearly describe the problem to be solved or the solution that is considered.

There are numerous books and websites available about how to gather evidence in a valid and reliable way – some of these aspects were discussed in previous modules. For most organizations traditional quantitative and qualitative methods will suffice. Quantitative methods a e used to gather data (evidence) that can be quantified, that is, measured and written down with numbers. We usually obtain these data from surveys or structured interviews. However, when it comes to exploring or identifying stakeholders' feelings and perceptions, qualitative methods are often more appropriate.

Qualitative methods are often exploratory in nature: they are used to gain a better understanding of underlying feelings, opinions, or motivations. One of the most widely used qualitative methods to gather evidence from stakeholders are focus groups: a set of six to ten people who are asked about their perceptions, feelings, or opinions regarding an assumed problem or proposed solution. The questions are asked by a skilled moderator in an interactive group setting where participants exchange point of views with other group members.

Another widely used qualitative method are unstructured or semi-structured interviews. Unstructured interviews don't use a predetermined questionnaire and may simply start with an opening question such as 'Can you tell me how you feel about X?' and then advance depending on the stakeholder's response. When interviewing stakeholders, it is important to inform them in advance about why you need their input.

Another widely used qualitative method are unstructured or semi-structured interviews. Unstructured interviews don't use a predetermined questionnaire and may simply start with an opening question such as 'Can you tell me how you feel about X?' and then advance depending on the stakeholder's response. When interviewing stakeholders, it is important to inform them in advance about why you need their input.

You also need to set clear expectations regarding anonymity, confidentiality, and how their information will be used, as this increases the likelihood of honesty.

Podcast: ACQUIRE Evidence from Stakeholders

In this podcast host Karen Plum discusses with Eric Barends, Managing Director, Center for Evidence-Based Management , Denise Rousseau, H J Heinz University Professor, Carnegie Mellon University, Dr Lisa J Griffiths, CEO, OzChild National Support Office, and Steven ten Have, Partner TEN HAVE Change Management & Professor of Strategy and Change at VU University Amsterdam, evidence from stakeholders – the people affected by our decision or the people who are able to influence it. The appraisal of this evidence is covered in the next module and its corresponding podcast episode.

In this podcast we explore the nature and extent of stakeholder individuals and groups – from governments and regulators, to organisational shareholders, the people implementing the decision and those that will live with the consequences. Rather than assuming that we know how these groups will react to our plans or solutions, we have to consult them.

In the experience of the guests, there is much to be learned which will inform and enrich the decision making process – including gaining an understanding of the complexities of implementing solutions and unearthing consequences we might not have anticipated.

Stakeholder engagement, assessment and management are all critical aspects of successfully implementing change – and understanding their points of resistance can help to unblock obstacles or just make the process smoother and the outcome more effective.

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

References

1. Pieters, J (2017). "*KLM Cabin Crew Announce 24 Hour Strike at Schiphol Next Month*." NL Times. nltimes.nl/2017/12/04/klm-cabin-crew-announce-24-hour-strike-schiphol-next-month.
2. Butler, S & Topham, G. (2017). "*Uber stripped of London Licence due to Lack of Coporate Responsibility.*" The Guardian. www.theguardian.com/technology/2017/sep/22/uber-licence-transport-for-london-tfl
3. Magness, V (2008). "Who are the stakeholders now? An empirical examination of the Mitchell, Agle, and Wood theory of stakeholder salience." *Journal of Business Ethics.*
4. Reynolds, S. J. & Heckman, D. R. (2006). "Stakeholder theory and managerial decision-making: Constraints and implications of balancing stakeholder interests." *Journal of Business Ethics*.
5. Popper, N (2016). "I*n Settlement's Fine Print, Goldman May Save $1 Billion*." The New York Times, Business & Policy.
6. The United States Department of Justice (2016). https://www.justice.gov/opa/pr/goldman-sachs-agrees-pay-more-5-billion-connection-its-sale-residential-mortgage-backed.
7. These three examples are adapted from "10 Great Examples of Ethical Decision Making In Business" by Barr D. Moore, *Great Work Life*, May 7, 2020. Retrieved from https://www.greatworklife.com/ethical-decision-making-in-business-examples/ on August 5, 2020
8. Nut, P. C. (2008). "Methods of data collection in qualitative research: interviews and focus groups." *British Dental Journal*.
9. Britten, N (1999). *Qualitative interviews in healthcare*. BMJ Books. 11-19 Pages.
10. Oltmann,S (2016). "*Qualitative interviews: a methodological discussion of the interviewer and respondent contexts.*" Forum: Qualitative Social Research.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 10.1

a: Correct

b,c,d: Incorrect

In the first scenario, the evidence f om stakeholders (the faculty members) indeed concerns stakeholder evidence (their experience that the faculty's management does not recognise them for their research publications). It concerns subjective feelings and perceptions, indicating that the dean's decision (introducing financial incentives) most likely won t have an effect.

However, in the second scenario, the evidence from stakeholders (the workers) concerns information regarding objectifiable facts (int oduction of a new pouring technique) related to the problem (increased number of honeycombs). Thus this does not concern stakeholder evidence, but rather organizational evidence.

Did I get this? 10.1

b: Correct

a,c,d: Incorrect

In the second scenario, the evidence from stakeholders (the workers) concerns stakeholder evidence (their judgment that the manager's decision is unfair and increases their workload). It concerns subjective feelings and perceptions, indicating that the managers' decision (required training) may not be accepted.

However, in the first scenario, the evidence f om stakeholders (the faculty members) concerns information regarding objectifiable facts (the outcome of a meta-analysis egarding the effect of a financial bonus). Thus this does not conce n stakeholder evidence, but rather scientific evidence

Learn by doing 10.2

a: These groups are entities within the organisation, so they are internal stakeholders.

b: The students' parents are not directly impacted by the decision, so they are indirect stakeholders.

c: The school does not have a direct responsibility towards this group, so this is a secondary stakeholder.

Did I get this? 10.2

	Internal	External
Direct	The coordinating manager	The local community
Indirect	The company's shareholders	The shipping company

The coordinating manager >

You could argue that the coordinating manager will be directly affected by the board's decision (she may even loose her job), which makes her a direct stakeholder. However, the company's management is an entity within the organisation, which makes her a direct internal stakeholder.

The local community >

The local community is an entity outside the organization, which makes them an external stakeholder. In addition, they will be directly affected by the board's decision, which makes them a direct external stakeholder.

The shipping company >

The shipping company is an entity outside the organisation, which makes them an external stakeholder. In addition, they will only be indirectly affected by the board's decision (eventually their revenues may decrease when the contract is terminated), which makes them an indirect external stakeholder.

The company's shareholders >

You may argue that the company's shareholders is an entity within the organisation, which makes them an internal stakeholder. However, they will only be indirectly affected by the board's decision (eventually their shares may drop in value), which makes them an indirect internal stakeholder.

Learn by doing 10.3

INFLUENCE / INTEREST	High	Some	Little
High	The shareholders	The local community	The coordinating manager
Some	The regulating authority	The legal department	The shipping company
Little	The local press	Resellers	Employees

a) The shipping company:

Some/Little > The decision may have limited impact on the shipping company (they may lose their contract, but that should not be insurmountable), and they have very limited power to influence th decision.

b) The regulating authority:

Some/High > The regulating authority has (some) interest in the decision, as their primary task is to make sure company's abide by the environmental laws. In addition, their power to influence th decision is high, because they can impose economic sanctions on the company.

c) Resellers:

Little/Some > Resellers have very little interest in the decision, and they have only limited influence t influence the decision (they could th eaten to choose another supplier).

d) The legal department:

Some/Some > The legal department has some interest in the decision: the legal requirements for the disposal of chemical waste could create extra work them. In addition, they have some power to influence the decision, as their expertise and cooperation is needed by the boa d.

e) The coordinating manager:

High/Little > The potential impact of the company's decision is high (she may lose her job), but she has almost no power to influence the decision

f) The local press:

Little/High > The coverage of the decision by the local press could create negative media attention that may harm the company's reputation, so their influence is high. Howeve , they are not impacted by the company's decision, so their interest is low.

g) Employees:

Little/Little > The company's employees have only little interest in the decision: they are from the local community, so there may be a small chance that some people hold the company's decision against them. In addition, they have almost no power to influence the decision

h) The shareholders:

High/High > The shareholders' interests are high: when the company's reputation is damaged its market value will drop, and as a result the shareholders will lose money. In addition, shareholders can fi e the board of directors, so their power to influence the decision is high

i) The local community:

High/Some > We would argue that the interests of the local community is very high (their environment and health could be affected), but that they have limited power to influence the company s decision.

Did I get this? 10.3

1: H – The Shareholders

2: B – The Regulating Authority

3: F – The Local Press

4: I – The Local Community

5: D – The Legal Department

6: C – The Resellers

7: E – The Coordinating Manager

8: A – The Shipping Company

9: G – The Employees

You don't agree? You may be right. Continue with the section 'Ethical Considerations' and the next Learn by doing exercise (10.4)

Learn by doing 10.4

We would consider the local community to be the most relevant stakeholder, and rank order the shareholders way lower.

However, keep in mind that whether a stakeholder is considered relevant is also determined by ideology, moral principles, and a company's corporate culture. For example, some companies feel that customers, employees, and the society at large are their most important stakeholders, whereas other (often listed) companies consider corporate profit and the inte ests of shareholders to be the most important. As a result, a business decision can be considered by some people as unethical, whereas others consider it 'normal business practice'.

Learn by doing 10.5

We would argue that the shareholders, the HR department and the managers have power to directly influence the decision or its implementation. This means that you should ask yourself: What a e the interests and concerns of these stakeholders and how much influence do they have – do they have the power to block the decision or impede its implementation? If their power to influence is high, you should acquire evidence about these shareholders' feelings and perceptions of the decision.

Learn by doing 10.6

We would argue that the employees and their families – and ultimately the society at large – all benefit from this decision. But also the shareholders, as the decision apparently has a positive effect on the company's sales – and consequently its market value.

Did I get this? 10.4

We would argue that all stakeholders benefit f om the company's decision to reduce its environmental impact. The only group we could think of that may experience harm are the company's competitors.

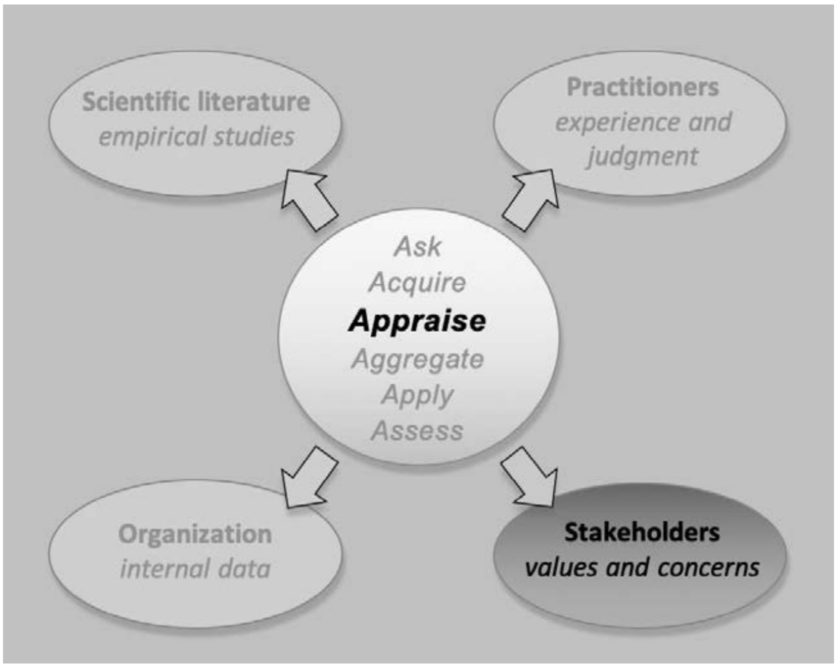
MODULE 11

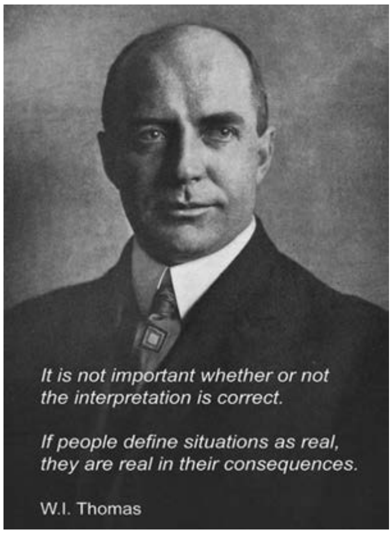
APPRAISE Evidence from Stakeholders

Learning objective:

- Assessing whether evidence from stakeholders is representative of the population from which it is drawn.

Introduction





Stakeholders are people whose interests affect or are affected by an organization's decision or its outcomes. As explained in the previous module, not all evidence *from* stakeholders is *stakeholder evidence*. Sometimes evidence from stakeholders actually represents organizational data or practitioner judgment, for instance when the evidence stakeholders provide involves actual facts regarding the assumed problem or the appropriateness of a preferred solution. Stakeholders can also provide evidence of their own subjective feelings and perceptions, for example, whether a problem or solution might affect their interests.

In the previous module, we explained why it is important to distinguish the types of evidence stakeholders can provide. With practitioner-, organizational-, and scientific evidence, we can critically appraise their reliability and validity. In contrast, with stakeholder evidence – feelings and perceptions – reliability and validity are not the major issue, but rather whether the evidence accurately represents the feelings and perceptions of all stakeholders.

This means that we appraise evidence from stakeholders differently when it concerns actual facts, that is, as organizational evidence or professional judgment. The present module on the other hand focusses on the critical appraisal of evidence regarding stakeholders' perceptions and feelings, core elements of stakeholder evidence. Where stakeholder evidence is concerned, *impact* and *representativeness* are the major focus of critical appraisal.

The importance of subjective feelings and perceptions

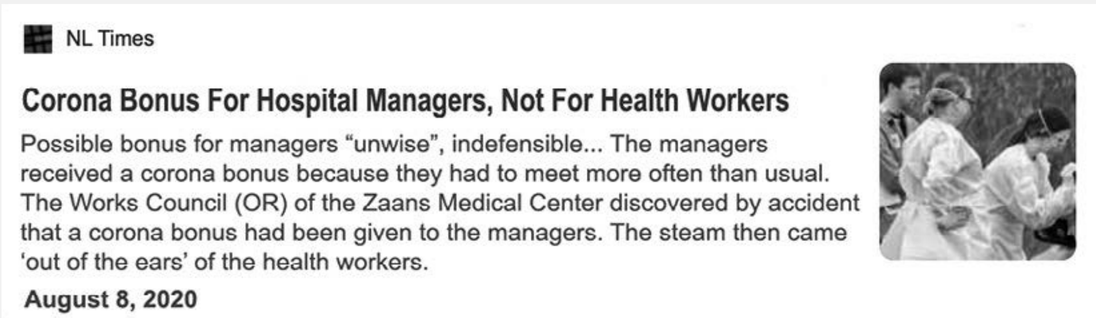
You may wonder why stakeholders' perceptions and feelings – evidence that is highly subjective and sometimes seemingly irrational – have a place in evidence-based decision-making. The answer is in the opening quote of this module. It reflects what is known as the Thomas Theorem — it basically means that what people feel and perceive to be true constitutes a social fact that has an actual effect on them – whether these feelings and perceptions are true or not. (1 (2

How stakeholders perceive a decision is a social fact too, regardless of whether their perception is based on subjective feelings, irrational beliefs, or personal values. As illustrated by the KLM and Uber examples in the previous module, sometimes stakeholders have the power to block a decision or impede its implementation. Taking into account stakeholder evidence can therefore substantially increase the likelihood of a favorable outcome of that decision. It helps decision makers identify ways to reduce avoidable harms or obtain the informed consent of those stakeholders who incur the risks of a negative outcome from that decision.

In addition, attention to stakeholders contributes to the ethical nature of decisions. Ethical deliberation can improve decision quality, by prompting decision-makers to consider options with better outcomes for all parties involved. Because of the broader information considered, using evidence regarding stakeholder perspectives and feelings tends to improve decision quality as well as the short- and long-term outcomes of your decision. (3) (4)

Learn by doing 11.1

Consider the following newspaper article



NL Times

Corona Bonus For Hospital Managers, Not For Health Workers

Possible bonus for managers "unwise", indefensible... The managers received a corona bonus because they had to meet more often than usual. The Works Council (OR) of the Zaans Medical Center discovered by accident that a corona bonus had been given to the managers. The steam then came 'out of the ears' of the health workers.

August 8, 2020

"The health workers of the ZMC Hospital got very angry about a 'corona bonus' that was given to managers. The hospital's Works Council coincidentally heard about the bonuses, which were given to the managers because – due to the COVID-19 outbreak – they had to make extra hours.

The Works Council believes that if managers are entitled to a bonus, everyone is entitled to a bonus. "How do you break the morale of your staff who have worked so very hard in recent months? Really bizarre," one person concerned told the newspaper.

The CEO explained that the managers received the bonus because it was not possible to compensate them for the extra hours they made. With the CEO wanted to treat all employees the same. The CEO regrets the uproar and understands the emotional reaction of the health workers, but points out that all staff – including managers, coordinators and team heads – have worked very hard during the COVID crisis and that paying a bonus for extra hours worked is only fair. " [5]

Should the CEO take the health workers' emotions and concerns regarding the board's decision into account?

a) No, because the CEO is right: the hospital should treat all employees the same, so if the managers were not compensated for extra work hours, paying them a bonus is only fair.

b) Yes, because even though the health workers' reactions are highly subjective and maybe even irrational, it has an actual impact on them.

c) No, because the health workers' reactions are highly emotional and even irrational – if the CEO and the board would reconsidering their decision this would seriously damage their authority.

d) Yes, because the health workers' reaction may lead to negative media coverage which could damage the hospital's reputation.

What is the impact of the decision?

The module '*APPRAISE – Evidence from the Scientific Literature*' explains that when critically appraising evidence from the scientific literature, we first need to determine whether the findings are of practical relevance.

To determine this, we have to look at the impact of the findings. Before we critically appraise the trustworthiness of stakeholder evidence, i.e. its representativeness, here too we must first determine the practical relevance of the evidence by looking at the impact. In the previous module, we have explained that we can distinguish two types of impact: 1) the extent of influence a stakeholder has on a decision, and 2) the extent of harm a decision brings on a stakeholder.

Impact: Practical relevance

The first type of impact concerns practical relevance: Evidence indicating that a powerful and influential g oup of stakeholders perceive a decision as unfavorable. Unfavorable perceptions on the part of important stakeholders can have serious practical implications for the decision-making process, particularly when these stakeholders have the power to block a decision or impede its implementation. In the previous module, it is explained how you can use a power/interest diagram to determine a stakeholder's impact.

Impact: Ethical relevance

The second type of impact concerns ethical relevance: Evidence indicating that a specific g rup of stakeholders may be harmed in some fashion by a decision. Potential harms have serious ethical implications for the decision-making process, particularly when they are perceived as unfair. A restructuring program in which only employees of 50 years of age and older are laid off is likely to be perceived as unethical by many people. Ethical concerns, however, not only arise from unevenly distributed harms, but also from unevenly distributed benefits. Imagine a company that has developed a self-driving car that decreases the number of fatal accidents involving pedestrians by 90%. Now imagine that the remaining 10% tend to be Black people. Although this self-driving car saves lives, the fact that it tends to save the lives of non-Black people is a serious ethical concern.

Learn by doing 11.2

Consider the following scenario

In 2015, the Home Office, the British government department responsible for immigration and passports, has introduced a new system to streamline visa applications. The system takes the information provided by visa applicants and automatically processed it, giving each person a colour code based on a "traffic light" system – green, amber, or red. The code red indicates an increased risk, so applicants with this code receive more intensive scrutiny by Home Office officials, take longer to determine, and are more likely to be refused. Contrastingly, applicants with code green receive less scrutiny and move swiftly through the process.

When software engineers are asked by Foxglove, an activist group of lawyers and technology experts, about how the system works, they explain that it is based on a computer algorithm that uses several metrics, one of which is nationality. The Home Office uses a list of "suspect nationalities" – if an applicant's nationality is on the list, this person would automatically be given a red rating.

Based on this finding, Foxglove argued that the system amounted to racial discrimination and characterized it as "speedy boarding for white people".

The Home Office, however, said it did not accept that description. "The system is just a streaming tool that has substantially improved the speed and efficiency of the visa application process."

1. **What is the impact of the system on the stakeholders – the visa applicants, in particular those that are on the list of "suspect nationalities"?**
   a) There is mainly a practical impact
   b) There is mainly an ethical impact
   c) There is both a practical and an ethical impact
   d) There is no impact

A senior civil servant thinks that this negative impact on the visa applicants won't change the decision to use the system, because they cannot be considered an influential group of stakeholders that has the power to stop the Home Office from using it.

2. **Do you agree?**
   a) Yes, I agree
   b) No, I don't agree

Representativeness

Critical appraisal is the process of carefully and systematically assessing the evidence in order to judge its trustworthiness. Critical appraisal of scientific evidence, for example, looks at a study s methodological appropriateness and examines factors such as internal validity. Critical appraisal of practitioner evidence, on the other hand, evaluates the extent to which a practitioner's judgment could be affected by cognitive biases. Stakeholder evidence, however, concerns subjective feelings and perceptions that we don't necessarily consider as objectifiable facts. For this reason, representativeness is a major indicator of stakeholder evidence quality.

Representativeness means how well the data obtained regarding stakeholder perspectives accurately represents all stakeholders in a particular group or category. The more representative the sample, the more confident we can be that we can generalize the evidence to the whole population. For example, a focus group of six employees might be used to get a sense of how employees in general might react to a planned change. But if the change potentially affects hundreds of employees, this sample may be insufficiently representative. In that case we could improve representativeness by corroborating the focus group's results with a follow-up survey of a larger sample of employees.

In the previous module, several methods are described for gathering stakeholder evidence. The trustworthiness of the evidence that results from these methods depends on 1) the opportunity stakeholders had to freely express their views and feelings regarding the problem or decision and 2) the representativeness of the stakeholders included in the sample. The first aspect is largely determined by the skills of the moderator (in case of focus groups) or the interviewer (in case of interviews), and the way questions are worded. The second aspect – sample representativeness – is determined by the way that sample was obtained.

Learn by doing 11.3

In the module "ACQUIRE – Evidence from Practioners" we have discussed the importance of developing of questions that accurately measure peoples' experience or judgment. The same counts for assessing stakeholders' feelings and perceptions. Regardless of the method used, the outcome is useless if the evidence gathered results from ambiguous, emotionally laden, or leading questions.

Are the following three questions adequately formulated to assess the feelings and perceptions of stakeholders? Select Yes or No.

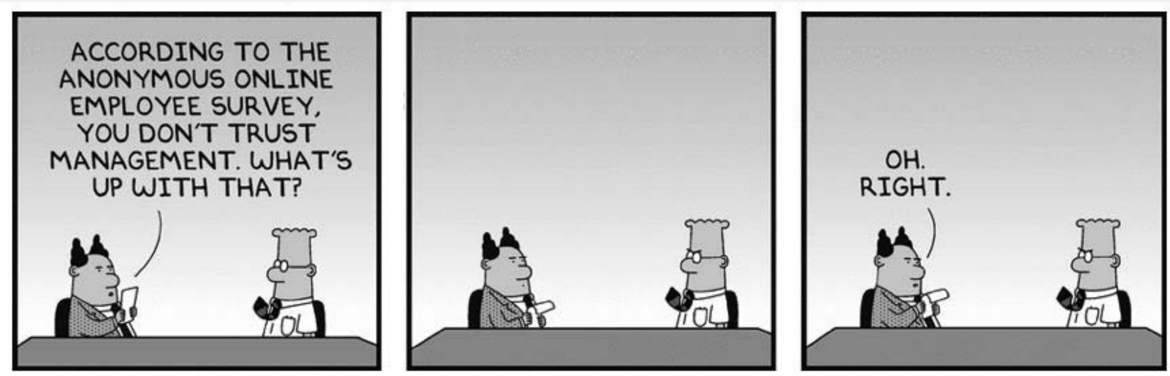
1. Do you agree that the introduction of the new accounting system will be a huge improvement compared to the old system?
2. Do you see any downsides or unintended negative consequences of the introduction of the new accounting system? Explain your answer.
3. Do you support the implementation of the new accounting system?

Learn by doing 11.4

Consider the following scenario

To assess stakeholders' feelings and perceptions towards an organisational change initiative that will affect their current way of working, a British telecom organization intents to send out a short questionnaire. The questionnaire contains some demographic questions such as postal code, date of birth, and job title along with five open, well formulated questions regarding their attitudes towards the proposed change.

Could there be a caveat of using the questionnaire to assess stakeholders' feelings and perceptions? Explain your answer.



How representative is my sample?

In the previous module, a power/interest diagram was provided to determine who – given the decision at hand – the most relevant stakeholders are. Obviously, the most representative data would not be a sample at all, but include all relevant stakeholders. In most cases, however, for practical reasons, you need to get the evidence from a smaller portion (sample) of the whole population of stakeholders. Then the challenge is to obtain a representative sample.

Although it sounds straightforward, obtaining a truly representative sample can be a challenge. For example, contrary to what is often assumed, the size of the sample has no direct relationship with representativeness; even a large random sample can be insufficiently representative. In fact, no sample is 100% representative.

As we discussed in the module '*ACQUIRE Evidence from Practitioners*', our main concern here is selection bias. Selection bias, also referred to as sampling bias, occurs when your selection of stakeholders leads to an outcome that is different from what you would have gotten if you had obtained evidence from the entire group of stakeholders. You can minimize selection bias by taking a random sample of the population. When a sample is randomly selected, each member of the population has an equal chance of being chosen. Thus, variation between the characteristics of the stakeholders in the sample and those of the entire group (population) is just a matter of chance.

Learn by doing 11.5

Consider the following scenario

A company decides to conduct interviews with employees to understand their feelings and perceptions regarding the proposed introduction of flexible working hours. On Friday at 9 a.m., the first 50 employees who start their working day are interviewed by trained interviewers.

Which of the following might negatively affect the outcome of the interviews?

a) A sample of 50 is not representative
b) Friday at 9 a.m. is not a good moment to administer a survey.
c) The sample is not random.

Unfortunately, even a random sample might, by chance, turn out to be anything but representative. For this reason, you should always (especially when a non-random sample is used) check whether your sample matches the characteristics of the entire group. You can do this by generating relevant base statistics for the entire population, and then comparing them with those of your sample. Examples of such base statistics are:

Age	**Occupation**
Gender	**Function**
Ethnicity	**Position**
Level of education	**Tenure**
Income	**Full-time/part-time**

When the characteristics of your sample are comparable to those of the entire population, you may safely assume that you have obtained a representative sample, and that you can generalize the evidence to the whole group of stakeholders.

Learn by doing

Consider the following scenario

The manager who is responsible for the interviews regarding the introduction of flexible hours decides to take a random sample of 50 employees. He asks the HR department to draw a random sample from the company's personnel system. He then compares the base statistics of the sample with those of the entire population. Below the results are presented.

Variable	Sample *n = 50*	Organization *n = 5,672*
Gender	-	-
Male	65%	47%
Female	35%	53%
Age *(mean)*	39	41
Level of education	-	-
Technical/vocational training	5%	7%
Associate degree	42%	47%
Bachelor's degree	32%	28%
Master's degree	21%	18%
Marital status	-	-
Single	22%	25%
Married or domestic partnership	78%	75%
Children	-	-
Yes	84%	78%
No	16%	22%
Tenure *(mean years)*	10	12
Employment	-	-
Part-time	38%	17%
Full-time	62%	83%
Household income	-	-
Less than $25,000	17%	13%
$25,000 – $50,000	21%	23%
$50,000 – $100,000	52%	46%
$100,000 – $200,000	8%	10%
More than $200,000	2%	8%

Do you consider the sample of employees to be a good representation of all the company's employees?

a) Yes

b) No

We started by explaining why stakeholders' feelings and perceptions – evidence that is highly subjective and sometimes seemingly irrational – have a place in evidence-based decision-making. What stakeholders feel and perceive to be true constitutes a social fact that has an actual effect on them, and as such can affect the outcome of a decision, in particular when stakeholders have the power to block a decision or impede its implementation.

Next, we discussed that before we critically appraise the trustworthiness of stakeholder evidence, i.e. its representativeness, we must first determine the practical relevance of the evidence by looking at the impact. We have explained that we should consider two types of impact: 1) Practical impact: Evidence indicating that stakeholders perceive a decision as unfavorable or having negative consequences; and 2) Ethical impact: Evidence indicating that a *specific* group of stakeholders may be harmed in some fashion by a decision.

We then discussed a major indicator of stakeholder evidence quality: representativeness. Representativeness means how well the data obtained regarding stakeholder perspectives accurately represents all stakeholders in a particular group or category. The more representative the sample, the more confident we can be that we can generalize the evidence to the whole population. The representativeness of stakeholder evidence depends on 1) the opportunity stakeholders had to freely express their views and feelings regarding the problem or decision and 2) the representativeness of the stakeholders included in the sample.

Finally, we discussed how to determine whether evidence from a sample of stakeholders is representative for all the stakeholders in the organization. Our main concern here is selection bias. You can minimize selection bias by taking a random sample of the population. Unfortunately, even a random sample might, by chance, turn out to be anything but representative. For this reason, you should always check whether your sample matches the characteristics of the entire group.

Podcast: APPRAISE Evidence from Stakeholders

In this podcast host Karen Plum discusses with Eric Barends, Managing Director, Center for Evidence-Based Management , Denise Rousseau, H J Heinz University Professor, Carnegie Mellon University, Dr Lisa J Griffiths, CEO, OzChild National Support Office, and Steven ten Have, Partner TEN HAVE Change Management & Professor of Strategy and Change at VU University Amsterdam, evidence from stakeholders – the people affected by our decision or the people who are able to influence it. The acquisition of this evidence is covered in Module 10 and its corresponding podcast episode.

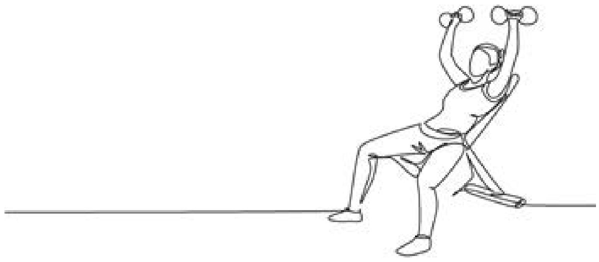
In this podcast we are reminded that sometimes we engage with certain parties as practitioners (those with experience in the problem or solution we're working with) and sometimes as stakeholders – with feelings and perceptions about how our decision will impact them, and how they might want to affect it, if they have sufficient power.

We hear several examples from the guests about stakeholder situations which show how things can be quite complicated, with lots of things at stake. We consider a number of ethical considerations, because at the end of the day, even though it may be possible to make a decision and implement a solution, it is also necessary to ask – should we?

Stakeholder evidence has the potential to throw us off course, perhaps slowing down our timetable and complicating matters. But consider the consequences if we don't know their perspectives and concerns or if we haven't weighed up whether their views are representative. There may be ethical, political, organisational and societal outcomes that could do irreparable harm.

https://evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 10.1: Stakeholders – retrospective

Think about a management, business, or policy decision you have been involved in making (or have observed closely). This decision should be reasonably important for the organization, involve significant resources and several or many people, and not have an 'easy' answer. Write down your answers to the following four questions in detail.

1. What exactly was the problem to be solved (or opportunity to be addressed)?
2. Who were the stakeholders? Make an overview in which you classify each stakeholder as an internal, external, direct, indirect, primary, or secondary stakeholder.
3. Who were the most relevant stakeholders? Make a table that maps the stakeholder's influence (power to exert influence over the decision and/or its outcome) and interests (perception of harms and benefits)
4. a. How was the evidence from each stakeholder group obtained?

b. How would you judge the reliability?

c. What could have been done to increase the reliability of stakeholder evidence?

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

Exercise 11.1: Stakeholders – prospective

Take a management, business, or policy decision you are involved in making (or are able to observe closely). The decision should be reasonably important for the organization, involve significant resources and several or many people, and have no 'easy' answer. Write down your answers to the following three questions in detail.

1. What exactly is the problem to be solved (or opportunity to be addressed)?
2. Who are the most relevant stakeholders? Make a table that maps the stakeholders' influence (power to exert influence over the decision and/or its outcome) and interests (perception of harms and benefits)
3. What would you consider to be the best method to obtain evidence from these stakeholders? Describe (in detail) your approach, including possible questions you would ask.

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

Suggestions for further reading

- *Stakeholder theory and managerial decision-making: Constraints and implications of balancing stakeholder interests*; Reynolds, Journal of Business Ethics, 2006
- *Ethics in Focus Groups: A Few Concerns*; Smith, QHR, 1995
- *Methods of data collection in qualitative research: interviews and focus groups*; Gill, BDJ, 2008
- *Managing Stakeholders for the Sake of Business and Society*; Fifka, New perspectives on corporate social responsibility, 2015
- *Focus Groups - CEBMa Info Sheet*; CEBMa, 2018
- *Interviews and focus groups in qualitative research: an update for the digital age*; Gill, BDJ, 2018

References

1. Thomas, W.I. & Thomas, D. S. (1928). *The Child in America: Behavior Problems and Programs*. Knopf.
2. Merton, R. K. (1995). "*The Thomas Theorem and the Matthew effect.*" Social Forces.
3. Hagafors, R. & Brehmer, B (1983). "Does having to justify one's judgements change the nature of the judgement process?." *Organizational Behavior and Human Performance*.
4. Nutt, P. C. (2002). *Why decisions fail.* Barrett-Kohler.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 11.1

a,c: Incorrect

b,d: Correct

Even though paying managers a bonus may be fair, the way this decision is perceived has an actual impact on the health workers and may consequently negatively affect their organizational commitment, work satisfaction, and performance. In addition, we would argue, there is another reason to take the health workers' emotions and perceptions into account: their reactions may lead to negative media coverage which could damage the hospital's reputation.

Learn by doing 11.2

1a,b,d: Incorrect

1c: Correct

We would argue that the practical impact on these visa applicants is high: they receive more intensive scrutiny, the application process takes longer, and they are more likely to be refused. In addition, the ethical impact is also high: the visa applicants are treated differently – and thus discriminated – based on their nationality.

2a: Incorrect

2b: Correct

We agree with the civil servant's observation that the visa applicants cannot be considered an influential g oup of stakeholders that has the power to stop the Home Office f om using the system. However, there are other stakeholders that are more powerful, such as the media, activist groups, and the society at large. In fact, after the news about the discriminating algorithm came out in 2020, the system was widely condemned and the Home Office decided to stop using the system

Learn by doing 11.3

1: No > This is a leading question that contains emotionally laden words ('huge improvement'). Formulating the question in this way may lead to social desirability bias. Better would be to ask an open question, for example: "*How do you feel about the introduction of the new accounting system?*"

2: Yes > The words that are used in this question are clear and unambiguous, so this is an adequately formulated question.

3: No > Although the words that are used in this question are clear and unambiguous, respondents will be inclined to give a socially desirable answer (especially when their anonymity is not guaranteed).

Learn by doing 11.4

We would argue that a possible caveat could be the use of demographic questions such as postal code, date of birth, and job title. The stakeholders may feel that these data could make it possible to identify the person behind the answers. Given the sensitive nature of the questionnaire's topic, they may therefore decide not to participate or give socially desirable answers.

Learn by doing 11.5

a: Incorrect > The size of a sample has no direct relationship with representativeness; even a large random sample can be insufficiently epresentative.

b: Correct > Employees who start their working day on Friday at 9 a.m. may differ from employees who start their working day at 10 a.m. or even later. Also, some employees with a part-time contract may have a regular day off on Friday. This means that these two groups of workers are likely to be underrepresented in the outcome of the interviews.

c: Not quite right. You are correct that interviewing first 50 employees who start their working day on Friday 9 a.m. is not a random sample, but there is a better answer.

Learn by doing 11.6

Yes: Incorrect

No: Correct

We would argue that in the sample there is an over-representation of male employees and part-time workers. This may affect their attitudes towards flexible working hours

CEBMa center for Evidence-Based Management

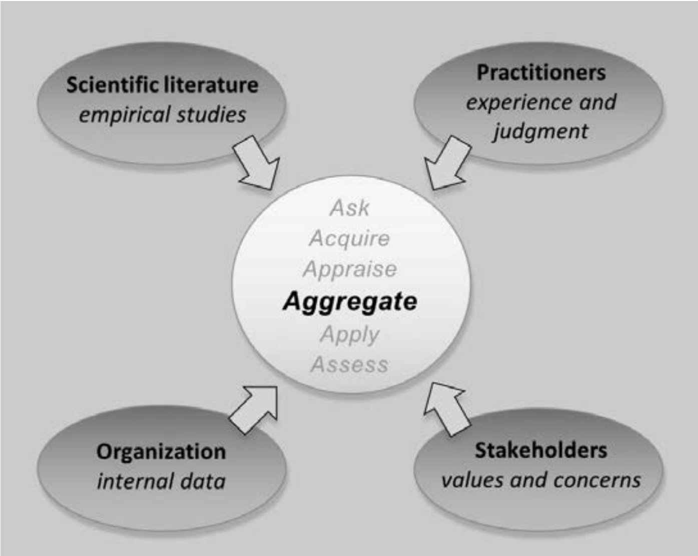
Open Learning Initiative

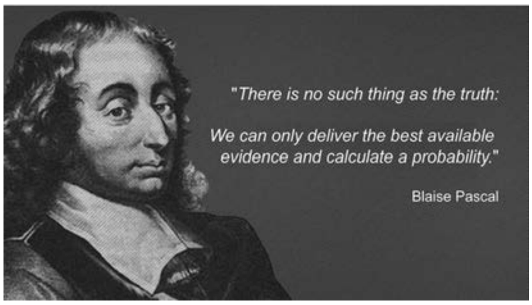
MODULE 12 | AGGREGATE: Weigh and Pull Together the Evidence

Learning objective:

- Apply Bayes rule to determine whether the available evidence supports a claim, assumption, or hypothesis.

Introduction





Example

Imagine the following situation

The board members of a large financial service firm meet to discuss the company's poor performance – for the fourth year in a row, the firm performs below the average for its sector, and none of the interventions to improve this situation have worked. The firm's HR director points out to his colleagues – who are all white, 50- to 60-year-old men with MBAs – that companies with a more diverse workforce tend to perform better. To substantiate his claim, he refers to a report, "Why Diversity Matters," [1] by McKinsey & Company, America's largest and most prestigious consulting firm. This report examines the data of 366 public companies across a range of industries in Canada, Latin America, the United Kingdom, and the United States. The findings of the report are clear: companies in the top quartile for ethnic diversity are 30 percent more likely to have above-average financial returns for their respective industries.

Taking an evidence-based approach, the board asks 10 experienced professionals within the company whether they support the claim that investing in an ethnically diverse workforce will lead to substantially better financial performance, with an increase of at least 10 percent. Most professionals state that, based on their experience at work, they believe that this claim is likely to be true.
Next, the scientific literature is consulted. A comprehensive search of the scientific literature yields five meta-analyses representing more than 150 studies that all demonstrate very small (and sometimes even negative) correlations between ethnic diversity and financial performance.

The organizational evidence shows a similar picture. There seems to be no difference in financial performance between the teams and departments with ethnically diverse workforces and those that have a more homogeneous makeup.

However, a sample of eight of the most important stakeholders, including regulators and institutional clients, indicate that they too believe increased ethnic diversity will have a substantial impact on the company's financial performance.

The company's CEO now sees himself faced with a difficult problem. He and his colleagues have taken an evidence-based approach, but now the evidence seems to be far from equivocal and even contradictory in some ways. So, what should he decide? Should he assign more weight to the evidence from the practitioners and stakeholders? They all seem to be very confident, but the CEO knows that human judgment, even f om experienced professionals, is often flawed. And what about the five meta-analyses? They are all based on cross-sectional studies (not longitudinal or controlled research), although findings all point in the same direction.

The same goes for the evidence from within the organization where higher-performing units do not tend to have a more diverse workforce. And, finally, what about the McKinsey report? Surely America's largest and most prestigious consulting firm cannot be wrong – or can they?

The CEO in the example above is faced with several challenges. First, he must weigh the different sources of evidence. As explained in the module The Basic Principles of Evidence-Based Management, not all evidence is created equal – some may count more than others. But how should the CEO balance the evidence from different sources, especially when they contradict each other? Second, how can he combine the evidence into one overall probability score? Finally, how can he make a decision based on this probability score?

In this module, we demonstrate how you can conduct the fourth step of evidence-based management: aggregate – weigh and combine evidence from different sources. Aggregating evidence, however, requires some understanding of probabilistic thinking and what is referred to as the Bayes rule. Although this may sound somewhat daunting, you probably, without realizing it, already apply Bayes rule when making daily decisions.

The first step to probabilistic thinking and Bayesian inference, however, is to understand the difference between a probability estimate and the truth.

Evidence-based management is about probabilities

Truth and proof

You may think that the purpose of an evidence-based approach is to find out whether a claim, assumption, or hypothesis is true or false. This is not the case. First, evidence is not the same as data. Where data can be numbers or figu es that can exist on their own, evidence exists only in the context of a claim or an assumption. Simply put, evidence is always evidence for (or against) something. Hence, data become evidence only when they stand in a *testing* relationship with a claim or a hypothesis.

Second, in evidence-based management, the term ***evidence*** is used deliberately instead of proof. This word choice emphasizes that evidence is not the same as proof and that evidence can be so weak that it is hardly convincing or so strong that no one doubts its correctness. In fact, proof is a concept that is useful only in the realm of mathematics. You can create a proof indicating that a mathematical statement or equation is true, but in many other domains, such as the social sciences, you cannot create proof of anything. This leads us to our third point, which is the concept of *truth*. Just as evidence is not proof, the outcome of an evidence-based process is not the *truth* but rather an *estimate of a probability*.

Truth – like proof – is a concept from a different domain. In mathematics, you can prove that an equation is true, but you cannot prove a claim or hypothesis in the domain of management. In fact, you could argue that in management, as in the social sciences generally, there is no such thing as *the truth*. After all, findings f om empirical studies are often influenced by multiple variables, and thus you can never definitely prove causality.

Instead, as the 17th-century French scientist and philosopher Blaise Pascal noted, we can only deliver the *best available evidence* and calculate a *probability*. And this is exactly what evidence-based management is about: making decisions under conditions of uncertainty through the use of the best available evidence from multiple sources to increase the probability of a favorable outcome. This means that the question for the CEO in our example is not whether it is true that an ethnically diverse workforce will lead to a higher financial performance, but rather what the *probability* is, *given the available evidence*, that an ethnically diverse workforce will lead to a higher financial performance

Did I get this 12.1

Read the following news article fragment

Wednesday could bring more fireworks than those being ignited for Fourth of July celebrations. Physicists at CERN's Large Hadron Collider, the world's biggest atom smasher, intend to announce tomorrow that they have proven the existence of the Higgs boson, also known as the God particle. Proof of the Higgs boson would be a massive milestone in the quest to understand "the most basic nature of the universe."

Is the use of the term *proof* in the article correct or incorrect?

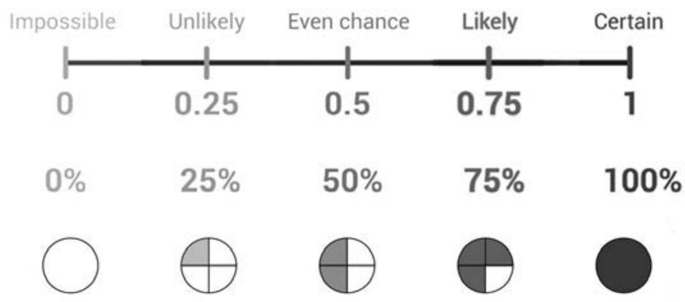
a) Correct, because physics is an exact science, so the use of the word proof is justified

b) Incorrect, because although physics is an exact science, the word proof is inaccurate

Probability versus chance

In evidence-based management, we refrain from using the words *proof* and *truth*; instead, we use terms such as *probable*, *likely*, *chance*, and *odds*. All these terms are related to the concept of probability, which is the extent to which something is likely to happen or to be the case. In daily life, probability is the same as chance, but in the realm of science (and evidence-based management), it is a calculation of the chance of an event taking place in percentage terms. Thus, probability is a ***measure*** of the chance that an event (or outcome) will occur. Probability, indicated with the letter *P*, is quantified as a number between 0 and 1, so P = 0 indicates impossibility (a chance of 0 percent that an event will happen), and P = 1 indicates an absolute certainty (a chance of 100 percent that an event will happen).

The higher the probability of an event (or outcome), the more certainty that the event will occur. In our example, if the probability of the HR director's claim is high, then it is more certain that investing in an ethnically diverse workforce will result in higher financial performance. If this probability is 0.8, it means that the certainty is 80 percent. If this probability is 0.4, it means that the certainty is only 40 percent, and thus there is a 60 percent chance that a diverse workforce will *not* lead to a higher financial performance or that it may even result in a lower financial performance. When the probability is 0.5, the certainty is 50/50, like flipping a coin, meaning that both a positive and a negative effect on financial performance are equally likely.



Learn by doing 12.1

A drawer contains 10 socks: 4 red socks and 6 blue socks. I randomly pick one sock. Which of the following two statements is correct:

1. There is a 40 percent chance that the sock is red.

2. There is a probability of 0.6 that the sock is blue.

a) Statement 1 is correct.

b) Statement 2 is correct.

c) Statements 1 and 2 are both correct.

d) Statements 1 and 2 are both incorrect.

Did I get this 12.2

Read the following scenario

The chief operating officer of a small soap manufacturing company suggests playing music in the production facilities, because listening to music while working makes the workers more productive. After consulting the research literature on this topic, it is estimated that the probability that the claim is true is 0.35.

Which of the following two statements is correct:

1. There is a 65% chance that listening to music will not lead to higher productivity.

2. Listening to music while working may harm productivity.

a) Statement 1 is correct.
b) Statement 2 is correct.
c) Statements 1 and 2 are both correct.
d) Statements 1 and 2 are both incorrect.

Why probability is always conditional

Imagine that someone asks you what the probability is that it will be raining tomorrow. (2) Your answer to this question would depend on several factors, such as the country you are in, the time of the year, and how the weather is today. Your estimate of the probability of rain tomorrow is therefore conditional on the information (evidence) you have available. If you know that you are in Holland during the autumn and that today it is raining heavily, your estimate of the probability will be different from the situation in which you know that it is summer, that the current weather is beautiful, and that the forecast says it will be sunny tomorrow. This also means that when the available evidence changes, the probability of it raining may do so too. It follows that to deal with probability in the realm of evidence-based management, it is necessary to recognize that the probability of a claim, assumption, or hypothesis being true is always conditional on the available evidence.

Notation

At this point it is necessary to introduce some notation. We already know that the probability is denoted with the letter *P*. For the claim or hypothesis of concern, we use the letter *H*. Thus, we denote the probability of the hypothesis being true as $P(H_{true})$. As explained previously, the probability P of the hypothesis H is conditional on the available evidence, so we use the letter *E* for evidence. This leaves us with the notation of “conditional on,” which is denoted as a vertical line.

$$\mathbf{P(H_{true}|E)}$$

Thus, the notation $\mathbf{P(H_{true}|E)}$ represents the probability of the hypothesis being true *given the available evidence*.

Learn by doing 12.2

Consider the following claim

According to a recent meta-analysis based on a large number of controlled studies, sales training improves the sales performance of young, inexperienced sales agents by 10 to 15 percent.

What is your estimate of the probability that this claim is true, given the available evidence?

a) **$P(H_{true}|E) < 0.5$.** The chance that this claim is true, *given the available evidence*, is lower than 50 percent.
b) **$P(H_{true}|E) > 0.5$.** The chance that this claim is true, *given the available evidence*, is higher than 50 percent.
c) **$P(H_{true}|E) = 0.5$.** The chance that this claim is true, *given the available evidence*, is 50 percent.

Learn by doing 12.3

Consider the following claim

According to a famous Harvard professor, sales training improves the sales performance of young, inexperienced sales agents by 10 to 15 percent.

What is your estimate of the probability that this claim is true, given the available evidence?

a) **$P(H_{true}|E) > 0.5$.** The chance that this claim is true, *given the available evidence*, is higher than 50 percent.
b) **$P(H_{true}|E) = 0.5$.** The chance that this claim is true, *given the available evidence*, is 50 percent.
c) **$P(H_{true}|E) < 0.5$.** The chance that this claim is true, *given the available evidence*, is lower than 50 percent.

Did I get this 12.3

Consider the following claim

In the 15 years I have worked as a sales manager, I have learned that sales training improves the sales performance of young, unexperienced sales agents by 10 to 15 percent.

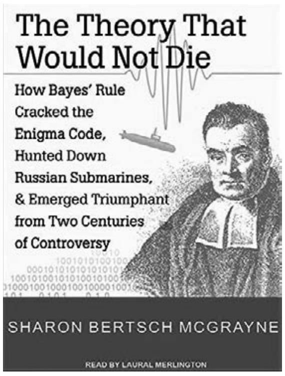
What is your estimate of the probability that this claim is true, given the available evidence?

a) **$P(H_{true}|E) > 0.5$.** The chance that this claim is true, *given the available evidence*, is higher than 50 percent.
b) **$P(H_{true}|E) = 0.5$.** The chance that this claim is true, *given the available evidence*, is 50 percent.
c) **$P(H_{true}|E) < 0.5$.** The chance that this claim is true, *given the available evidence*, is lower than 50 percent.

Bayes rule

As explained previously, the purpose of the fourth step of evidence-based management – aggregate – is to answer the question, *What is the probability of the hypothesis (claim, assumption) being true given all the available evidence?* We can write this as $P(H_{true}|E) =$. To find an answer, we need the help of Reverend Thomas Bayes, an English minister and amateur mathematician who lived between 1701 and 1761. Very little is known about Bayes's life, even though he lent his name to a whole new branch of statistics and to a hugely influential theorem. At some point during the 1740s, Bayes came up with his rule:

The probability of a hypothesis being true given the evidence depends on both prior knowledge and the likelihood of the evidence.



Then, in 1774, the mathematician Pierre-Simon Laplace (independently of Bayes's work) formulated the same rule in its current form: that of a mathematical equation. After more than two centuries of controversy, Bayes rule is now widely applied in fields such as genetics, image processing, epidemiology, forensic science, and medical diagnostics. The strength of Bayes rule is that it can deal with all kinds of evidence, making it applicable to all kinds of questions that involve probability. For example, during World War II, Bayes rule was used to crack the Enigma code, and during the Cold War, it helped to find a missing H-bomb and to hunt down Russian submarines. It has also been used to investigate the safety of nuclear power plants, predict the tragedy of the space shuttle Challenger, demonstrate that smoking causes lung cancer, build Google's search engine, and much, much more. (3)

On the face of it, the Bayes rule is a simple one-line theorem: Posterior probability equals prior probability times likelihood. Mathematically, this can be written as follows:

$$P(H|E) = \frac{P(H)\,P(E|H)}{P(E)}$$

If you are put off by this equation, don't worry: we have developed a downloadable app for your smartphone that will do the calculation for you. (4) In addition, you can use the online calculator on the CEBMa website. (5) The instructions to download and use the app or the calculator are provided in the learning activities.

Bayes rule can be written mathematically in many different forms. Although each form is somewhat different, all contain the same three elements: the prior probability, the likelihood of the evidence, and the posterior probability.

1. **The prior probability:**

   $P(H_{true})$ = The initial estimate of how probable it is that the hypothesis (claim, assumption) is true to start with.

2. **The likelihood of the evidence:**

   $P(E|H_{true})$ = The likelihood of the evidence being available/showing up if the hypothesis were *true*.

   $P(E|H_{false})$ = The likelihood of the evidence being available/showing up if the hypothesis were *false*.

3. **The posterior probability:**

   $P(H_{true}|E)$ = The revised estimate of the probability of the hypothesis being true given the available evidence.

In the following sections, we discuss each element in more detail and demonstrate how you can use Bayes rule to combine evidence from different sources and calculate an overall probability.

The prior probability

$$P\ (H_{true})$$

The prior probability (known simply as "the prior") is the initial estimate of how probable it is that the claim or hypothesis is true to start with, that is, without the benefit of the available evidence. In most cases, we should set this prior probability at 0.50, which means we do not have any reason to assume the hypothesis is either false or true. This is the case when we know nothing about the hypothesis or have no prior relevant experience that may help us to determine its probability. We may have a strong opinion, but in general, merely having an opinion is not sufficient to set a reliable prior. This may change, however, when *baseline information* is available. Consider the following example:

Example

Marie-Claire is 23 years old, her favorite writer is Marcel Proust, and her preferred holiday destination is France. In her leisure time, she loves to read French poetry. Which of the following is more likely?

A. Marie-Claire is a law student.

B. Marie-Claire studies French literature.

Although you may be inclined to answer B, the right answer is actually A. There are many more law students than French literature students – so many more, in fact, that a student who reads French poetry and likes spending her holiday in France is more likely to be a law student.

Sometimes a prior probability can be very strong and have a large impact on the overall probability. This is not to suggest that prior probabilities always dominate the available evidence and thus the outcome of Bayes rule. In fact, in the realm of management, the opposite is often true: in many cases, there is no reliable prior available other than your professional judgment. So, unless we have a reliable prior estimate available, such as the average in the sector, incidence/prevalence numbers, general statistics, or census data, we set the prior at P = 0.50 (meaning 50/50, like flipping a coin)

Learn by doing 12.4

Read the following scenario

At the airport of Toronto, you come across a tall, burly, hockey-stick wielding man wearing a Maple Leafs shirt, the logo of the Toronto Maple Leafs, a professional ice hockey team based in Toronto.

What is more likely?

1. **The man is a tourist.**
2. **The man is a professional ice hockey player for the Toronto Maple Leafs.**

a) Option 1: The man is a tourist.
b) Option 2: The man is a professional ice hockey player for the Toronto Maple Leafs.
c) Both options are equally likely.

Prior probability = context

As explained above, the prior is the baseline probability that a claim, assumption or hypothesis is true. It provides the context in which the probability of claim being true should be judged. For example, when a dead body that is found is claimed to be Mr. Johnson, this claim is more likely to be true when the body was found in Mr. Johnson's house then when it was found in a forest 100 miles away. In this case the context – the location where the body was found – functions as a prior that sets the baseline for the probability that the body will be identified as Mr. Johnson.

In the domain of management, contextual information such as census data, average in the sector, incidence/prevalence numbers, and general statistics can provide us with a prior probability that can serve as a starting point to determine whether a claim or hypothesis is likely to be true.

Learn by doing 12.5

Read the following scenario

During a board meeting of a medium sized North American company, the HR manager claims that the number of employees taking sick leave has significantly increased in the past 6 months. The HR manager has not (yet) provided evidence for this claim. However, according to a recent report of the US Department of Labor the number of adult workers that were absent due to sick leave in this period has increased by 15 percent.

What is your estimate of the probability that the HR manager's claim is true to begin with?

a) Higher than 50 percent
b) About 50 percent
c) Lower than 50 percent

The likelihood of the evidence

In evidence-based management, the phrase "the likelihood of the evidence" is often used. This is shorthand for ***the likelihood of the evidence being available given the hypothesis, claim, or assumption***. The notion of likelihood takes into account two different aspects that can be considered opposite sides of the same coin. First, it estimates the likelihood of the evidence being available or showing up if the hypothesis were true: $P(E|H_{true})$. It then estimates the likelihood of the evidence being available or showing up if the hypothesis were false: $P(E|H_{false})$.

Probability versus likelihood

You will notice that we use the term probability when referring to the hypothesis and the term likelihood when referring to the evidence. In daily life, these terms are synonyms, but in statistics and probability theory, there is a distinction between them. The technical explanation for this difference, however, is beyond the scope of this course.

Using the online Bayes calculator

In the following sections, you will need to do some calculations. But don't worry: when you use our online Bayes calculator, you only have to type in some numbers and press CALCULATE. You can find the calculator on the website of the Center for Evidence-Based Management. We will provide you the link in the learning activities.

Let's apply prior probability and likelihood to a real-world example

To get a good understanding of the concept of likelihood, consider this example (6): *John comes home from a short business trip. When his wife unpacks his suitcase, she finds another woman's underwear.* Given the evidence – women's underwear in his suitcase – what is the probability that John has cheated on his wife? As mentioned, the likelihood of the evidence is determined by two elements:

1. **$P(E|H_{true})$:** *The likelihood of the evidence showing up if the hypothesis were true.*

   In this example, this is the likelihood of the underwear showing up in the suitcase if John were cheating on his wife. If so, it is easy to imagine how the underwear got there. Then again, even (and perhaps especially) if he were cheating on his wife, you might expect him to be extremely careful, particularly when he knew his wife might unpack her suitcase. On the other hand, when we discuss this example in class, female students often point out that most men are stupid, so the evidence may very well show up if the hypothesis were true. Nevertheless, we would argue that the probability of the underwear appearing in John's suitcase if he were cheating on his wife would be rather low – say 30 percent.

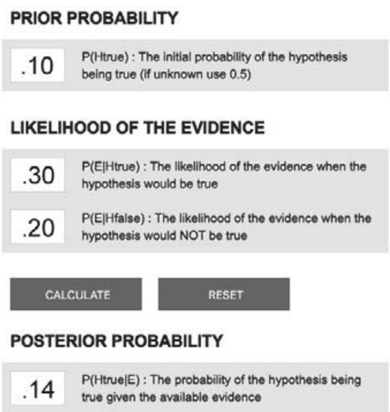
2. **$P(E|H_{false})$:** *The likelihood of the evidence showing up if the hypothesis were false.*

   In this example, this is the likelihood of the underwear showing up in the suitcase if John had NOT cheated on his wife. Could there be other plausible explanations for the garment being in John's suitcase? With some imagination, you can easily come up with several alternative explanations: maybe John's colleagues played a prank on him, or the contents of John's suitcase got mixed up during a security check at the airport, or the underwear might be a gift to his wife that he forgot to wrap. In fact, they could even be his! Taking all these less or more plausible alternative explanations into account, we would set this likelihood at 20 percent.

As you can see, the difference between $P(E|H_{true})$ and $P(E|H_{false})$ is rather small (0.3 versus 0.2), which indicates that the evidence is not very convincing. This means that, in this example, the likelihood of the evidence is rather low.

Now let's assume that the prior probability of John's cheating on his wife was very low to begin with – for instance, because John is a very faithful, honest person – so we set the prior at 10 percent. Use the online calculator to work out $P(H_{true}|E)$.

Link to online calculator: ***https://cebma.org/resources-and-tools/bayes/***



If you have typed in the values correctly, you will find a probability of only 0.14, meaning that the available evidence has increased the posterior probability by only 4 percent (from 10 percent to 14 percent). The reason for this slight increase is the low likelihood of the evidence: $P(E|H_{true})$ is only slightly higher than $P(E|H_{false})$.

Learn by doing 12.6

In some cases, the likelihood of the evidence is so powerful that it strongly updates the prior probability. Consider this scenario:

Your colleague claims that disabled employees tend to achieve higher performance ratings at work. Since there is no reliable prior probability available, other than your personal judgment, you set the prior at 50 percent. A search in ABI/INFORM (see Module 5), however, yields a well-conducted meta-analysis based on 13 randomized controlled trials (RTCs) demonstrating a large positive effect.

What is the probability of your colleague's claim being true, given the likelihood of the evidence?

- **$P(E|H_{true})$**
  This is the likelihood that a meta-analysis based on 13 RCTs would demonstrate a positive effect if your colleague's claim were true. If you have completed Module 7, APPRAISE Evidence from the Scientific Literature, you will agree with us that if the claim – disabled employees tend to achieve higher performance ratings at work – were indeed true, a meta-analysis of 13 RCTs would most likely show a positive effect, so we would set this likelihood at 90 percent.
- **$P(E|H_{false})$**
  This is the likelihood that a meta-analysis based on 13 RCTs would demonstrate a large positive effect if your colleague's claim were false. We would argue that if your colleague's claim were not true, it would be very unlikely that a meta-analysis would show a positive effect, so we would set this likelihood very low, at only 5 percent.

Apply Bayes rule to calculate $P(H_{true}|E)$, the probability that the claim is true given the evidence. What posterior probability do you get?

Link to online calculator: *https://cebma.org/resources-and-tools/bayes/*

a) Lower than 0.5
b) 0.5
c) Between 0.5 and 0.9
d) Higher than 0.9

Did I get this 12.4

Consider the claim we used earlier in this module:

In the 15 years I have worked as a sales manager, I have learned that sales training improves the sales performance of young, inexperienced sales agents by at least 10 percent.

Because you recently read a study on the impact of management training, you know that training programs in general have only a limited effect, so you set the prior at 60 percent. The evidence that is available is the sales manager's 15 years of professional experience. After you have asked the manager some critical questions about this experience (how often she was involved in sales training, how the effect on sales performance was measured), you assess the likelihood of this evidence as follows:

- **$P(E|H_{true})$**
  The likelihood that a sales manager with 15 years of professional experience would claim that sales training improves performance when this claim is indeed true. From the critical questions, you have learned that the manager has been involved in more than 20 sales training sessions and that the impact on sales performance was measured in an objective and reliable way, so you set this likelihood at 70 percent.
- **$P(E|H_{false})$**
  The likelihood that a sales manager with 15 years of professional experience would claim that sales training improves performance when this claim is actually false. You consider this to be possible: you learned in Module 4 that practitioner experience is prone to all kinds of cognitive biases, so you set this likelihood at 40 percent.

Apply Bayes Rule to calculate $P(H_{true}|E)$, the probability that the claim is true given the evidence. What posterior probability do you get?

Link to online calculator: https://cebma.org/resources-and-tools/bayes/

a) Lower than 0.5
b) 0.5
c) Between 0.5 and 0.9
d) Higher than 0.9

Aside I: The likelihood ratio

$$\text{Likelihood Ratio} = \frac{E \mid P(H_{true})}{E \mid P(H_{false})} = \text{strength of the evidence}$$

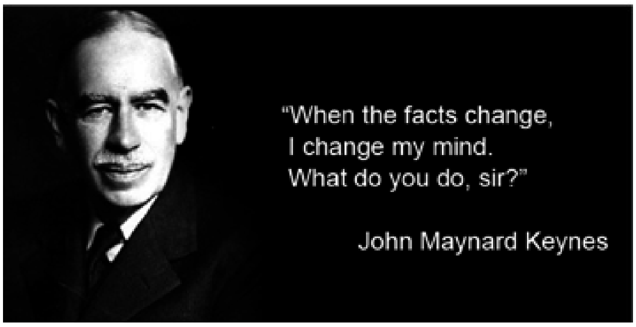
In the (slightly sexist) example of the women's underwear in the suitcase, the likelihood of the evidence in favor of the hypothesis was rather low. In the example of the disabled employees getting higher performance ratings, the likelihood of the evidence was very high. You may wonder whether there is a way to quantify the strength of the evidence. There is. It is referred to as the 'Likelihood Ratio'. This metric can be calculated by dividing $P(E|H_{true})$ *– the likelihood of the evidence supporting the hypothesis – by* $P(E|H_{false})$ *– the likelihood of the evidence contradicting the hypothesis. In the example of the women's underwear in the suitcase, the likelihood ratio is .30/.20 = 1.5, meaning that the evidence in support of the hypothesis that John cheated on his wife is only 1.5 times stronger than the evidence in support of the hypothesis that John did not cheat on his wife. In the example of the disabled employees, the likelihood ratio is .90/.005 = 180, meaning that the evidence in support of the hypothesis that disabled employees tend to get higher performance ratings is 180 times stronger than the evidence in support of the hypothesis that employees do not get higher performance ratings.*

Aside II: The prosecutor's fallacy

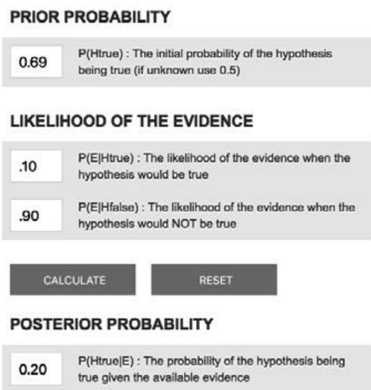
Many people assume that the likelihood of the evidence given the hypothesis P(E|H) is the same as the probability of the hypothesis given the evidence P(H|E). This is incorrect. Imagine, for example, an American court of law, where the forensic expert has demonstrated that there is a partial DNA match between the perpetrator and the defendant, and that only 1 in 100,000 persons have that same partial match. Based on this information, the prosecutor wrongly concludes that the probability of the defendant being innocent given this evidence is 1 in 100,000; thus, the probability of the defendant being guilty is 99.999 percent. This error is known as the "prosecutor's fallacy." After all, there are about 325 million people in the United States, meaning that 1 in 100,000 would still account for 3,250 partial matches, which would make the probability that the defendant would be guilty 1 in 3,250, which is 0.03 percent, not 99.999 percent.

Still confused? Think of it this way: some popes are Italian, but not all Italians are popes. Therefore, the probability of a person being Italian, given that he is a pope, is not the same as the probability that a person is a pope, given that he is Italian.

Updating the probability when new evidence becomes available



In the previous sections, we saw how we can use Bayes rule to calculate the probability of a claim or hypothesis being true given the available evidence. The strength of Bayes rule, however, is that it allows you to update this probability when new evidence comes available. Sometimes this new evidence is so strong that it can update the probability of a hypothesis from near zero to an almost certainty.



Consider this example from *The Signal and the Noise*. (6) Before the September 11 attacks, we would have assigned a near-zero probability to the possibility of terrorists crashing planes into a skyscraper in Manhattan. As you can read on the Wikipedia page ***September 11 Intelligence before the Attacks***, a fair educated guess would be 1 in 20,000, or 0.005 percent. However, we would also have assigned a very low probability to a plane hitting a skyscraper by accident. In fact, this probability can be estimated empirically: 1 in 12,500, or 0.008 percent. (7) When we use Bayes rule to calculate the posterior probability, we find that after the first plane hit the World Trade Center, the initial probability increased from 0.005 percent to 38 percent.

It is therefore not surprising that most people who witnessed the first plane hitting the World Trade Center believed that it was an accident. This, of course, changed when the second plane hit the World Trade Center.

The strength of Bayes rule is that it allows you to update the posterior probability when new evidence becomes available. In the example of the September 11 attacks, this means that we can update the posterior probability of a terror attack happening after the first plane hit the World Trade Center to 38 percent by using this as our prior probability when the second plane hit the World Trade Center. If you calculate the new posterior probability, you can see that, based on the new evidence (a second plane having hit the World Trade Center), the posterior probability of a terror attack becomes a near certainty: 99.99 percent.

The examples we have used so far demonstrate that we can apply Bayes rule to all types of evidence: medical tests, unfaithfulness, terror attacks, and findings f om social science. This means that we can use Bayes rule to combine evidence from different sources and estimate an overall probability.

Learn by doing 12.7

Consider the claim we used earlier:

In the 15 years I have worked as a sales manager, I have learned that sales training improves the sales performance of young, inexperienced sales agents by at least 10 percent.

Based on prior knowledge (a study on the impact of management training), you set the prior at 60 percent. Then you determined the likelihood of the evidence: 15 years of professional experience. Because the sales manager has been involved in more than 20 sales trainings, and the impact on sales performance was measured in an objective and reliable way, you might set the $P(E|H_{true})$ at 70 percent. But because practitioner experience is prone to all kinds of cognitive biases, you set the $P(E|H_{false})$ at 40 percent.

You then applied Bayes rule to calculate $P(H_{true}|E)$, and found that the probability that the claim is true, given the evidence, is **72 percent**.

Now let's assume that new evidence comes available: a meta-analysis of 20 randomized controlled trials (RCTs) demonstrating that the impact of sales training does indeed improve the sales performance of young, inexperienced sales agents by at least 10 percent.

Apply Bayes rule to calculate $P(H_{true}|E)$, the probability that the claim is true given the evidence. What posterior probability do you get?

Link to online calculator: ***https://cebma.org/resources-and-tools/bayes/***

a) Lower than 0.7
b) Between 0.7 and 0.8
c) Between 0.8 and 0.9
d) Higher than 0.9

HINT

$P(E|H_{true})$: If the claim were indeed true, a meta-analysis of 20 RCTs would very likely demonstrate a positive effect, so you should set this likelihood at **90 percent**.

$P(E|H_{false})$: If the claim were not true, it would be very unlikely that a meta-analysis of 20 RCTs would demonstrate a positive effect, so you should set this likelihood very low, at only **5 percent**.

Weighing and aggregating evidence from different sources

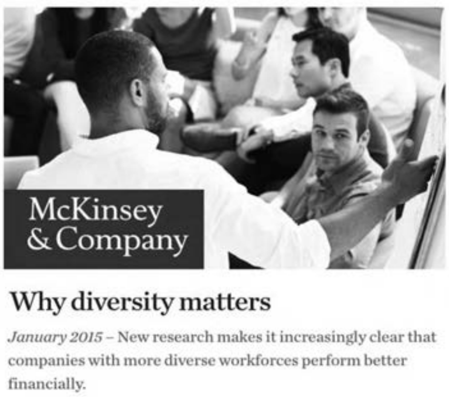
We started this module with an example of how evidence from multiple sources is sometimes contradictory. Even when the evidence is equivocal, however, the question remains how to combine it into one overall probability score. We can do this by applying Bayes rule. Bayes rule, however, works best with a claim or hypothesis that concerns a specific (e.g., numerical or dichotomous) outcome. Claims such as “X will substantially increase Y” are rather vague and are therefore hard to test against the evidence. In the case example cited at the start of this module, the claim contains a very precise outcome: a 10 percent increase in the company’s financial performance. We use this case as an example of how to aggregate and weigh evidence from multiple sources using Bayes rule.

Example

The question we have to answer is, ***What is the probability that the claim “an ethnically diverse workforce leads to an improvement in the firm’s financial performance by at least 10 percent”*** *is true, given the following evidence:*

a) ***Evidence from practitioners:*** *The judgment of 10 experienced professionals*
b) ***Evidence from the scientific literature:*** *Five meta-analyses representing more than 150 studies*
c) ***Evidence from stakeholders:*** *The opinions of the eight most important stakeholders*
d) ***Evidence from the organization:*** *The organizational data*

The prior probability



According to Bayes rule, we first need to determine a prior probability. In this case, we could use McKinsey’s report. So, what is the trustworthiness of this report? As you may recall from the previous modules, findings from a survey conducted by a commercial company such as McKinsey appearing in a nonacademic journal – for example, a magazine, report, or white paper – cannot be considered highly trustworthy. (9) In addition, a cross-sectional survey is not an appropriate research design for measuring the effect of ethnic diversity on financial performance. We therefore assign a low trustworthiness score to the report (55 percent) and a correspondingly low prior probability of 0.55.

How to assign trustworthiness scores to scientific evidence is explained in the module "*Appraise – Evidence from the Scientific Literature*"; a table for effect-studies can be downloaded << here >>

Source 1: Evidence from practitioners

The board obtained evidence from 10 experienced managers, but what is the likelihood of this evidence? To be more precise, what is the likelihood that experienced managers will agree that an ethnically diverse workforce will improve the company's performance by at least 10 percent if this hypothesis were true? We would argue that this likelihood is quite high, so we set $P(E|H_{true})$ at **90 percent**.

According to Bayes rule, however, we must also consider $P(E|H_{false})$: What is the likelihood that experienced professionals would agree if the hypothesis were NOT true (e.g., if the effect on performance were much lower than 10 percent, or even negative)?

We would argue that this likelihood is substantial because, as you learned in Module 4, human judgment, even from experienced professionals, is often flawed, especially in a situation where the work environment is rather unpredictable, and direct, objective feedback is lacking (as is the case here). We therefore estimate this likelihood at **50 percent**.

Learn by doing 12.8

Apply Bayes rule to calculate $P(H_{true}|E)$, the probability that the claim is true given the prior probability (the McKinsey report) and the likelihood of the evidence from practitioners. What posterior probability do you get?

Link to online calculator: ***https://cebma.org/resources-and-tools/bayes/***

a) Lower than 0.5
b) Between 0.5 and 0.7
c) Between 0.7 and 0.9
d) Higher than 0.9

HINT

The prior probability, **$P(H_{true})$**, is **0.55**.
The likelihood of the evidence, **$P(E|H_{true})$** and **$P(E|H_{false})$**, is **0.9** and **0.5**.

Source 2: Evidence from the scientific literature

Next, the board consulted the scientific literature and found five meta analyses representing more than 150 studies that all demonstrated very small and sometimes even negative correlations. What is the likelihood of this new evidence being available given the hypothesis?

Again, we must consider both $P(E|H_{true})$ and $P(E|H_{false})$. What is the likelihood that all five meta-analyses would demonstrate only small or negative correlations if the hypothesis – that an ethnically diverse workforce has a large, positive effect on performance – were true? We would argue that this would be *very unlikely*, so we set $P(E|H_{true})$ at **10 percent**.

We can apply the same logic to $P(E|H_{false})$: If an ethnically diverse workforce does NOT increase performance, then it would be *very likely* that five meta-analyses show only small or negative cor elations, so we set this likelihood at **90 percent.** (10)

Learn by doing 12.9

Apply Bayes rule to calculate P(H_{true}|E), the probability that the claim is true given the likelihood of the new evidence: five meta-analyses. What posterior probability do you get?

Link to online calculator: ***https://cebma.org/resources-and-tools/bayes/***

a) Lower than 0.5
b) Between 0.5 and 0.7
c) Between 0.7 and 0.9
d) Higher than 0.9

HINT

Don't forget to use the previous probability **(0.69)** as the new prior!

Source 3: Evidence from stakeholders

The board also consulted eight important stakeholders, who all stated that they too consider the hypothesis likely to be true. So again, what is the likelihood of this evidence? We would argue that P(E|H_{true}) and P(E|H_{false}) are similar to those of the experienced professionals, as the stakeholders' judgment is also prone to cognitive biases. (11)

Did I get this 12.5

Apply Bayes rule to calculate the new P(H_{true}|E), the probability that the claim is true given the likelihood of the new evidence: eight important stakeholders. As stated, the likelihood of this evidence is the same as the likelihood of the evidence from the practitioners (see above). What new posterior probability do you get?

Link to online calculator: ***https://cebma.org/resources-and-tools/bayes/***

a) Lower than 0.2
b) Between 0.2 and 0.4
c) Between 0.4 and 0.6
d) Higher than 0.6

Source 4: Evidence from the organization

Finally, the board consulted organizational evidence, which showed no significant difference in performance between the teams and departments with an ethnically diverse workforce and those with a more homogeneous makeup. Again, what is the likelihood of this evidence showing up given the hypothesis?

Learn by doing 12.10

We know this is a bit of a mind-bender, but give it a try. What is the likelihood that this organizational evidence would be found if the claim were true? Moreover, how would you estimate the likelihood that these organizational data would be found if the claim were NOT true?

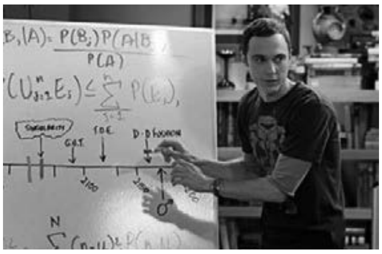
Did I get this 12.6

Apply Bayes rule to calculate the new P(Htrue|E), the probability that the claim is true given the likelihood of the new evidence: the organizational data. What posterior probability do you get?

Link to online calculator: ***https://cebma.org/resources-and-tools/bayes/***

a) Lower than 0.2
b) Between 0.2 and 0.4
c) Between 0.4 and 0.6
d) Higher than 0.6

So, the answer to our question is . . .



We have weighed and aggregated all four sources of evidence by using Bayes rule, so we can now answer our question:

What is the probability that the hypothesis that an ethnically diverse workforce leads to an improvement in the firm's performance by at least 10 percent is true, given the prior probability and evidence from 10 professionals, five meta-analyses, organizational data, and eight of the most important stakeholders?

The answer to this question is:

10 percent

This means there is a 90 percent chance that an ethnically diverse workforce will NOT lead to a performance improvement of 10 percent (that is, substantially lower or even a decrease).

The CEO in our example was initially faced with contradicting evidence: the professional and stakeholder evidence supported the HR director's claim that diversity will improve performance, whereas the scientific and organizational evidence did not. However, by using Bayes rule, we were able to weigh and aggregate the evidence and calculate an overall probability score (i.e., 10 percent). Based on this score, the CEO should conclude that investing in an ethnically diverse workforce is *not* the best way to increase the company's performance. Of course, there may be several reasons why a company should invest in a diverse workforce (e.g., for ethical or social reasons), but in this particular example, performance is not one of them.

Bayesian thinking

The strength of Bayes rule is that it allows us to aggregate different types of evidence – medical diagnoses, terror attacks, suspected cheating – and revise our initial estimate when new evidence becomes available. Bayes rule is not some kind of magic formula or merely a rule of thumb. It is a theorem, which means it has been mathematically proven to be true. (Note, this is the only place in the course where we claim something is proven to be true). But Bayes rule is more than a mathematical equation. It's a way of thinking that we can apply to all aspects of daily life. It can help us make better decisions in a world we can never fully understand. In addition, it makes us aware that claims, assumptions, and beliefs about how the world works are never black and white – true or false – but are grayscale. (12)

As pointed out earlier, most people already apply Bayes rule without realizing they are doing so: when we go through the world and encounter new ideas and insights, the level of confidence in our beliefs changes accordingly, especially when we face evidence that cannot be reconciled with prior beliefs. Since our beliefs are never 100 percent certain, our confidence in them changes when we encounter new evidence

Bayes rule also explains why some people are so hard to convince despite the evidence: if your prior assumption is close to zero, only an overwhelming amount of evidence can increase your prior belief. In fact, closed-minded people with a prior of zero will never learn anything from any evidence because anything multiplied by zero is still zero. (13)

In the same way, Bayes rule explains why Carl Sagan's dictum that *"extraordinary claims require extraordinary evidence"* is correct: the prior probability of an extraordinary claim is extraordinarily low (otherwise, it wouldn't be deemed extraordinary), so we need an extraordinarily high P(E|Htrue – and thus an extraordinarily low P(E|Hfalse – to move the needle. For this reason, any manager, leader, consultant, professor, or policymaker who makes an outlandish claim needs to provide more compelling, trustworthy evidence than those who make a more modest claim.

In this module, you have learned how to use Bayes rule to weigh and aggregate evidence from multiple sources. It is a formalization of how you can revise the initial probability of a claim or assumption being true when new or better evidence becomes available. Although it helps to understand the math behind Bayes rule, you don't need to learn the formula by heart – we have developed a smartphone app and online calculator to do that for you. What is more important than memorizing the formula by heart is to internalize the general idea behind it and learn to intuitively apply its basic principles: whenever a claim is being made, automatically consider the prior probability; estimate the likelihood of the evidence if the hypothesis or assumption is true, P(E|Htrue); estimate the likelihood of that same evidence if the hypothesis or assumption is false, P(E|Hfalse); and adjust your posterior probability when new, compelling evidence becomes available.

We started this module by explaining that evidence-based management is about probabilities, not silver bullets. For this reason, we refrain from using the words *proof* and *truth*, instead using terms such as *probability* and *likelihood*.

We explained that *probability*, indicated by the letter *P*, is quantified as a number between 0 and 1, so $P = 0$ indicates impossibility (a chance of 0 percent that an event will happen) and $P = 1$ indicates moral certainty (a chance of 100 percent that an event will happen). In addition, we explained that in the realm of evidence-based management, the probability of a claim, assumption, or hypothesis being true is always *conditional* upon the available evidence.

We then introduced Bayes rule: The probability of a hypothesis being true given the evidence depends on two elements: the *prior probability* and the *likelihood of the evidence*.

The *prior probability* (known simply as "the prior") is the initial estimate of how probable it is that the claim or hypothesis is true to start with, that is, without the benefit of the available evidence. In most cases, we should set this prior probability at 0.5, which means we do not have any reason to assume the hypothesis is either false or true. In some cases, however, a prior probability can be very strong and have a large impact on the overall probability. Mathematically, the prior is written as $P(H_{true})$.

The *likelihood of the evidence* takes into account two different aspects that can be considered to be opposite sides of the same coin. First, it estimates the likelihood of the evidence being available or showing up if the hypothesis were true: $P(E|H_{true})$. It then estimates the likelihood of the evidence being available or showing up if the hypothesis were false: $P(E|H_{false})$. If the difference between $P(E|H_{true})$ and $P(E|H_{false})$ is large, the likelihood of the evidence is high, indicating that the evidence is strong.

If you have estimated (or calculated) the prior probability and the likelihood of the evidence, you can use Bayes rule to calculate the probability of a claim or hypothesis being true given the available evidence. The strength of Bayes rule, however, is that it allows you to *update* this probability when *new evidence* becomes available.

Next, we used a case example to demonstrate how Bayes rule can be applied when we are faced with contradictory evidence from multiple sources. The question we answered was, *What is the probability that the hypothesis "an ethnically diverse workforce leads to an improvement in the company's performance by at least 10 percent" is true?* Based on the prior probability, the initial answer was 55 percent, but after weighing and aggregating evidence from practitioners, the scientific literature, stakeholders, and the organization by using Bayes rule, we ended with a probability of only 10 percent.

We ended this module by explaining that you do not need to know the formula of Bayes rule by heart (we have an app or online calculator for that). What is more important is to internalize the general idea behind it and learn to intuitively apply its basic principles: whenever a claim is being made, consider the prior probability, estimate the likelihood of the evidence, and adjust your posterior probability when new, compelling evidence becomes available.

Podcast

In this podcast host Karen Plum discusses with Eric Barends – Managing Director, Center for Evidence-Based Management, Denise Rousseau – H J Heinz University Professor, Carnegie Mellon University, and Julia Galef – Co-founder of the Center for Applied Rationality, how we bring together the various sources of evidence that we've gathered.

The purpose of taking an evidence-based approach is to reduce uncertainty in our decision making, looking at likelihoods and probabilities to guide our thinking and discussions. The use of Bayes rule and Bayesian thinking are explored, so that we continue to protect ourselves from falling prey to bias (particularly confirmation bias), but that we consider alternative explanations for the evidence that we found – if our initial belief is either true or false.

The use of probabilities isn't something our brains take to easily, so there is some challenge inherent in this approach, but it is simply an extension of the overall evidence-based management approach, where we look at each type of evidence and consistently question whether it is trustworthy, robust and reliable. Once we reach the 'aggregate" stage, it's time to ask how likely is it that the claim or hypothesis we are investigating is true (or false).

https://evidencebasedmanagement.buzzsprout.com

Suggestions for further reading

- *The Theory That Would Not Die. How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, Sharon Bertsch, Talks at Google, YouTube, Aug 2011
- *The Signal and the Noise*, Nate Silver, YouTube, 2013
- *Think Rationally via Bayes' Rule* | Big Think, Julia Galef, YouTube, 2013
- *Belief, bias and Bayes*, Jon Butterworth, The Guardian, Sept 2014
- *On the Importance of Bayesian Thinking in Everyday Life*, Michal Oleszak, Towards Data Science, Dec 2021

References

1. Hunt, V., Layton, D. & Prince, S. (2015). *Why diversity matters.* McKinsey & Company. https://www.mckinsey.com/business-functions/organization/our-insights/why-diversity-matters
2. Evett, I. W. (1998). "Towards a uniform framework for reporting opinions in forensic science casework." *Science & Justice.* Volume 38. Number 3. 198-202. Example and some of the explanations are adapted from Evett's work.
3. A great overview of the history and application of Bayes's rule is provided in Sharon Bertsch McGrayne, T*he Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy* (Yale University Press, 2011).
4. The app is called Bayes Manager and is available through the Apple store and Google Play.
5. https://www.cebma.org/resources-and-tools/bayes.
6. Adapted from: Silver, N., & Chamberlain, M. (2012). *The signal and the noise: Why so many predictions fail-but some don't.* Penguin. New York.
7. See Silver & Chamberlain (2012), p. 247: "*In the previous 25,000 days of aviation over Manhattan prior to September 11, there had been two such accidents: one involving the Empire State Building in 1945 and another at 40 Wall Street in 1946. That would make the possibility of such an accident about 1 chance in 12,500 on any given day*".
8. Bernstein, P. L. (1996). *Against the gods, the remarkable story of risk.* John Wiley & Sons. New York.
9. The findings would be more trustworthy if they came from surveys conducted by scholarly researchers, who strive to be objective, and if it were published in a journal maintaining a high standard of quality, accuracy, and academic integrity. Research conducted by organizations and published in press releases, newspapers, or magazines could be more biased or influenced by the desire to obtain certain findings.
10. You may be confused by the fact that we estimate the probability (and thus trustworthiness) of the findings of a meta-analysis based on cross-sectional studies at such a high percentage. After all, a cross-sectional study is not an appropriate design to examine a causal relation. This is correct. As you saw in the modules on scientific evidence, a cross-sectional study lacks both a control group and before and after measurements, so even when a meaningful, relevant correlation is found, only one out of three criteria for causality is met. However, when little or no correlation is found (evidence of no effect), none of the three criteria for causality is met, which is a strong and more trustworthy indication that no causal relation exists. Think of it this way: when it snows, there may be enough snow for skiing tomorrow (possibility), but when it doesn't snow, there will certainly not be enough snow for skiing tomorrow (high certainty).
11. For example, there could be authority bias-their judgment may be influenced by the findings of the McKinsey report.
12. Galef, J. (2013). *Think rationally via Bayes' rule.* Big Think. https://www.youtube.com/watch?v=NEqHML98RgU.
13. Butterworth, J. (2014). "*Belief, bias and Bayes.*" The Guardian. https://www.theguardian.com/science/life-and-physics/2014/sep/28/belief-bias-and-bayes.

Learn by doing & Did I get this? Answers and Feedback

Did I get this? 12.1

a: Incorrect

b: Correct

Science is about gathering information and testing assumptions (hypotheses) in ways that allow us to estimate how likely it is that something is true. We can create proof that a mathematical statement or equation is true, but even in physics, we cannot create proof of anything. We can only calculate a probability.

Learn by doing 12.1

a,b,d: Incorrect

c: Correct

There is a 40 percent chance that the sock is red. However, this means that there is a 60 percent chance that the sock is blue. A probability of 0.6 also means there is a 60 percent chance that the sock is blue, so statements 1 and 2 are both correct.

Did I get this? 12.2

a,b,d: Incorrect

c: Correct

A probability of 0.35 that the COO's claim is true means that the certainty is only 35 percent. Thus, there is a 65 percent chance that listening to music while working will not lead to higher productivity OR that it may even result in lower productivity. Therefore, statements 1 and 2 are both correct.

Learn by doing 12.2

a,c: Incorrect

b: Correct

In this case, there is strong evidence available that supports the claim – a recent meta-analysis based on a large number of controlled studies. It depends on the quality of the evidence whether we should assign a probability of 0.8 or even higher. Thus, the chance that this claim is true, *given the available evidence*, is higher than 50 percent.

Learn by doing 12.3

a,c: Incorrect

b: Correct

The fact that the claim is made by a famous Harvard professor is irrelevant: it does not tell us what the evidence base for the claim is (professional experience, scientific esearch?). In fact, we do not even know what the professor's field of expertise is (organ transplantation, biochemistry?). This means that there is no evidence available that supports or contradicts this claim, so we must assign a probability of 0.5. Thus, the chance that this claim is true, *given that no evidence is available*, is 50 percent (like flipping a coin)

Did I get this? 12.3

a: Correct

b,c: Incorrect

In this case, there is evidence available that supports the claim: 15 years of professional experience. As you learned in Module 4, APPRAISE Evidence from Practitioners, 15 years of experience does not automatically constitute trustworthy evidence, but it does move the probability of the claim over the threshold of 50 percent. Thus, the chance that this claim is true, *given the available evidence*, is higher than 50 percent.

Learn by doing 12.4

a: Correct

b,c: Incorrect

Although the man looks like a hockey player, the prior probability that he is a professional hockey player for the Toronto Maple Leafs is extremely low. Only 19 professional hockey players play for the Toronto Maple Leafs, while the airport of Toronto welcomes an average of 46,000 international passengers per day. Thus, the probability that the man actually plays for the Maple Leafs is much smaller than the probability that the man is a tourist who bought a Maple Leafs shirt and a hockey stick as souvenirs.

Learn by doing 12.5

a: Correct

b,c: Incorrect

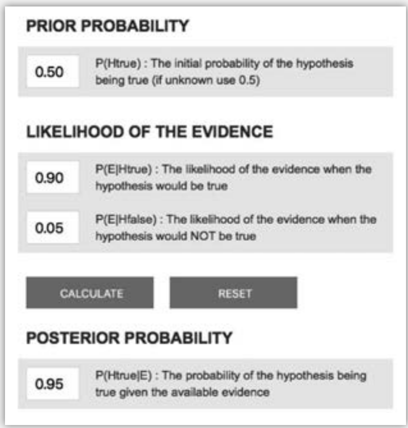
Although the HR manager has not yet provided any evidence, the information from the US Department of Labor serves as a prior, indicating that the probability that the claim is likely to be true is higher than 50 percent. Note that if this information is not available, the probability that the HR manager's claim is true drops to 50 percent.

Learn by doing 12.6

a,b,c: Incorrect

d: Correct

When you apply Bayes rule to calculate $P(H_{true}|E)$, you get a posterior probability of 0.95, meaning that due to the high likelihood of the evidence, the initial probability has increased from 50/50 (like flipping a coin) to 94 pe cent (a near certainty).

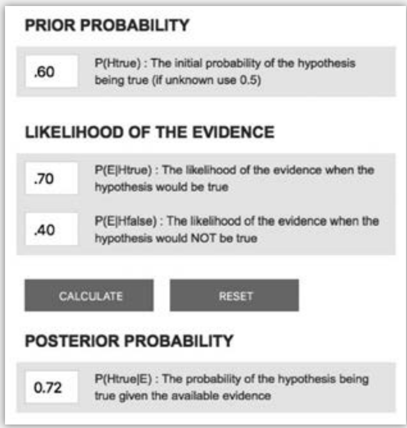


Did I get this? 12.4

a,b,c: Incorrect

d: Correct

When you apply Bayes rule to calculate P(Htrue|E), you get a posterior probability of 0.72, meaning that, due to the likelihood of the evidence, the probability that the claim is true has increased from 60 to 72 percent.

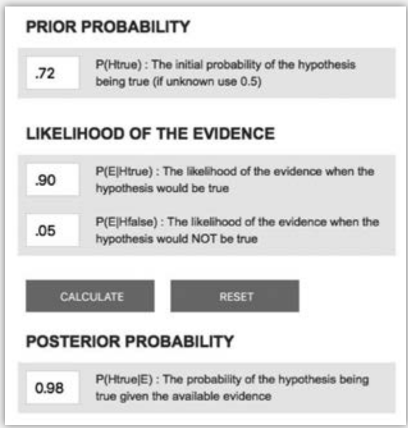


Learn by doing 12.7

a,b,c: Incorrect

d: Correct

When you use the initial probability that the claim is true as the new prior, you get a new probability of 0.98, meaning that due to the high likelihood of the new evidence, the initial probability has increased from 72 to 98 percent (an almost certainty).

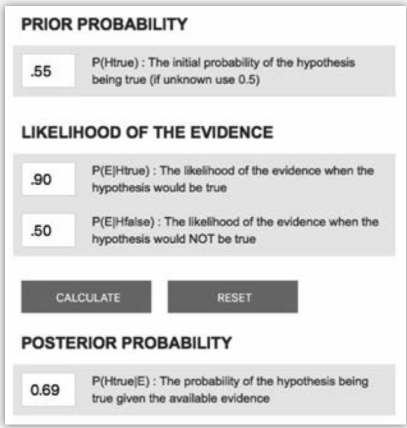


Learn by doing 12.8

a,c,d: Incorrect

b: Correct

When you apply Bayes rule you get a posterior probability of 0.69, meaning that due to the evidence of practitioners, the probability that the claim is true has increased from 55 to almost 70 percent.

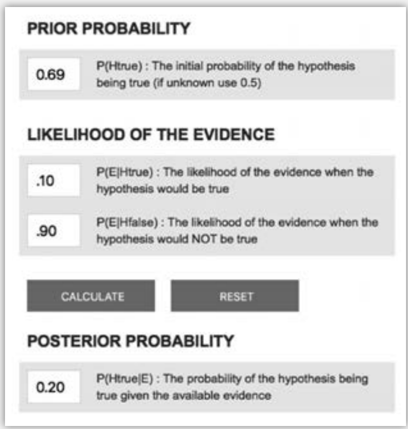


Learn by doing 12.9

a: Correct

b,c,d: Incorrect

When you apply Bayes rule, you get a posterior probability of 0.2, meaning that due to the evidence from the scientific literature, the probability that the claim is true has dramatically decreased from 69 percent to 20 percent.

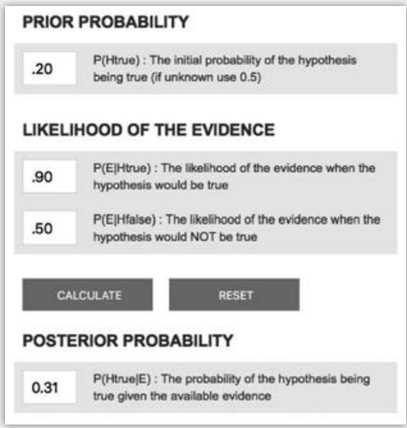


Did I get this? 12.5

a,c,d: Incorrect

b: Correct

When you apply Bayes rule, you get a posterior probability of 0.31, meaning that due to the evidence from the stakeholders, the probability that the claim is true has slightly increased from 20 percent to 31 percent.



Learn by doing 12.10

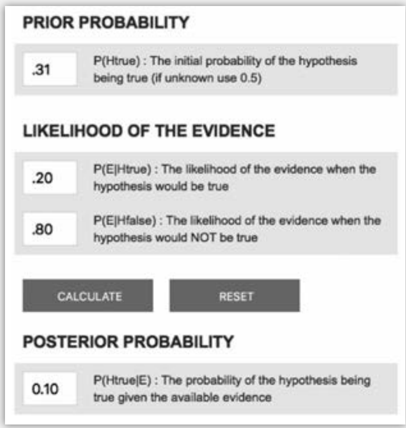
We would argue that the likelihood of the organizational data NOT demonstrating any difference if the hypothesis were true would be very low. Again, the opposite is true for $P(E|H_{false})$: if the hypothesis were false, we would expect the organizational data to show no difference (which they don't). We therefore estimate $P(E|H_{true})$ at 20 percent and $P(E|H_{false})$ at 80 percent.

Did I get this? 12.6

a: Correct

b,c,d: Incorrect

When you apply Bayes rule, you get a posterior probability of 0.1, meaning that due to the evidence from the organization, the probability that the claim is true has again decreased from 31 percent to 10 percent.



CEBMa center for Evidence-Based Management

Open Learning Initiative

MODULE 13 | APPLY: Incorporate Evidence into your Decision

Learning objectives:

- Determine whether the evidence applies to the organizational context.
- Determine a decision's costs and benefits.
- Determine whether the evidence is actionable.
- Determine whether moderators need to be taken into account.
- Determine, given the type of decision at hand, how and in what form the evidence can be applied.

Introduction

> *Knowing is not enough; we must apply.*
> *Willing is not enough; we must do.*
>
> – JOHANN WOLFGANG VON GOETHE

Example

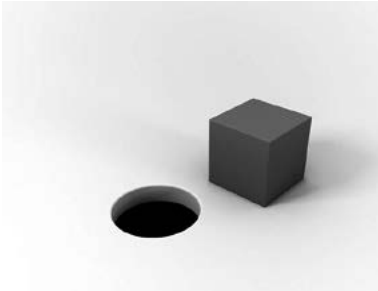
Consider the example of a board of directors in a large Canadian hospital.

The board is thinking about introducing a performance appraisal system that evaluates physicians' clinical performance and gives feedback to help them learn. After consulting the scientific literature, the board finds a large body of evidence suggesting that performance appraisal can indeed have positive effects on a person's professional development. These effects are, however, contingent upon a wide range of moderating factors, including the purpose of the appraisal, the raters' reliability, perceived fairness, ratee and rater personality, and so on (1).

The research literature also suggests that evaluations based on objective outcome measures tend to be more reliable than those based on personal judgments. In addition, the hospital's physicians report that they are more likely to accept feedback from other physicians (rather than from non-physicians), preferring feedback from respected colleagues. Overall, the board estimates that the probability that such a performance appraisal will enhance the physicians' professional development is about 70 percent. Given this evidence, should the board introduce the performance appraisal system?

In most cases, the answer to whether to implement a management practice is not a simple yes or no. Rather, *it depends*. It depends on whether the evidence applies to the particular organizational context and goals, whether the anticipated benefits outweigh the risks, and whether the evidence is *actionable*. In this module, we discuss the factors you need to take into account when applying evidence in the decision-making process.

Consideration 1: Does the evidence apply?



After we have acquired, appraised, and aggregated the evidence relevant to a problem or preferred solution, we must ask, Does the evidence apply? In other words, is the evidence generalizable to our organizational context? (2) This question is especially important when the evidence comes from people outside the organization or from external sources such as the research literature. The evidence may be valid and reliable, but it might come from a distinctly different industry, such as medicine, aviation, or the military. The circumstances in those industries may or may not be relevant to our organizational context. It is important to determine whether our context is different in relevant ways from the context in which the evidence was acquired. Contextual factors can influence the outcome of our decision or necessitate adaptations in how we apply the evidence. To determine whether the evidence applies to our organizational context, we can use the framework discussed in previous modules: the PICOC.

Population	*Who?*	Type of employee, people who may be affected by the outcome
Intervention	*What or how?*	Management technique/method, factor, independent variable
Comparison	*Compared to what?*	Alternative intervention or factor
Outcome	*What are you trying to accomplish, improve, change?*	Objective, purpose, goal, dependent variable
Context	*Organizational setting or circumstances*	Type of organization, sector; relevant contextual factors

In the example of a hospital considering whether to introduce a performance appraisal system, the PICOC could be formulated as follows:

P: Physicians

I: Performance appraisal

C: Status quo

O: Professional development

C: A large Canadian healthcare organization

The underlying notion is that all five PICOC elements can pertain to whether performance appraisal will have a positive effect on a person's professional development. Thus, the answer to the question "Does X work?" is less relevant than the response to the question "For whom does X work, for what purpose, and in which context?" Unfortunately, assessing whether the evidence is sufficiently generalizable to the organizational context is often subjective. No one-size-fits-all guideline exists for determining generalizability. Some research findings are generalizable to all human beings, but sometimes what works in one setting might not work in another. For example, there is strong evidence that a person's reaction to feedback determines the extent to which his or her performance will improve. How someone reacts to feedback, however, is partly determined by a person's openness to feedback. Some people argue that medical specialists – for example, surgeons – are less open to feedback than are other professionals, so for them, performance feedback may be less effective. However, resistance to feedback could also be a stereotype for which the evidence is limited. (3)

The same is true for stereotypes about the organization. As discussed in the module '*Basic Principles*', managers and leaders often assume that their organization is unique, making the applicability of external evidence appear limited. However, it is commonplace for organizations to have myths and stories about their own uniqueness. In general, they tend to be neither exactly alike nor unique, but somewhere in between.

Learn by doing 13.1

Read the scenario below

A financial services organization has experienced several incidents of unethical behavior. The HR director points out that high-quality studies in the realm of medicine have demonstrated that in healthcare organizations, role models – people we can identify with, who have qualities we would like to have, or who are in positions we would like to reach – have an important impact on the professional values, attitudes, and behavior of medical students and doctors. For this reason, he argues, it is important that the company's senior managers function as a role model: when they are perceived as ethical, employees are less likely to engage in unethical behavior.

Do you agree with the HR director that the findings of studies on the importance of role models (doctors) in healthcare organizations are also applicable to role models (senior managers) in a financial services organization?

a) Strongly agree
b) Agree
c) Neither agree nor disagree
d) Disagree
e) Strongly disagree

Learn by doing 13.2

Some studies in the social sciences – particularly randomized controlled trials – are conducted in artificial (laboratory-type) settings with students carrying out p escribed tasks. Obviously, the PICOC for these studies is very different than for a "real-world" organization: students are not employees (P), and a laboratory setting with researchers making observations is not a real work setting (C). Some people therefore argue that laboratory studies are good at telling whether or not some manipulation of an independent variable causes changes in the dependent variable, but that these results do not generalize to the real world.

What is your take on this?

Consideration 2: What is the expected value?

In the previous module, we explained how to estimate the probability of a claim or hypothesis being true regarding an assumed problem or preferred solution. We do so by weighing and aggregating the available evidence. By using Bayes rule, we can combine evidence from different sources and calculate an overall (posterior) probability to help us make an evidence-based decision. When there is only a 0.09 (9 percent) probability that a claim or hypothesis regarding a preferred solution is true, the conclusion is obvious: the available evidence suggests very strongly that the solution will not solve the problem. The decision in that case is clear: go back to the drawing board and come up with a better solution. But what if the probability score is somewhere between 0.3 and 0.6? In that case, what would be the best decision? The answer is that it depends on the *expected value*.

The decision as to whether or not to implement a solution depends on the notion of *expected value*: the (sometimes monetary) outcome expected from the decision. (4) To calculate the expected value, you need to know four things:

1. The cost-benefit of the solution if the claim is true (outcome 1
2. The probability that the claim is true given the evidence (P1)
3. The cost-benefit of the solution if the claim is false (outcome 2
4. The probability that the claim is false given the evidence (P2)

You can now calculate the expected value as follows:

Expected value = (P1)(outcome 1) – (P2)(outcome 2)

Example

Imagine that it is claimed that solution A will lead to a productivity increase of $1.5 million (outcome 1). Now let's assume the probability that this claim is correct, given the evidence, is 60 percent (P1). This means there is a 40 percent probability that this claim is incorrect (P2). Let's also assume that if the claim is false, the productivity may decrease by $150,000. In that case, the expected value is 900,000 (0.6 x 1,500,000) minus 60,000 (0.4 x 150,000) = $840,000. With this expected value, you would most likely decide to implement solution A.

Now consider the following example. Imagine that it is claimed that solution B will lead to a productivity increase of $300,000 and that the probability that this claim is correct, given the evidence, is again 60 percent. Now also assume that if the claim is incorrect, the productivity will decrease by $800,000. In this example, the expected value is 180,000 (0.6 x 300,000) minus 320,000 (0.4 x 800,000) = a $140,000 loss. With this expected value, you would most likely decide not to implement solution B. This demonstrates that your decision should depend not only on the probability a solution will work but also on the solution's anticipated costs and benefits.

Did I get this 13.1

Read the scenario below

A British food delivery company considers implementing a new smartphone app to communicate with its 450 delivery agents who deliver food from restaurants to its customers in Central London. It is expected that the app will lead to a faster delivery process (within 30 minutes), which will result in a productivity increase of $300,000. According to the company's IT director, there is an 85 percent chance that the app will work as planned. This, of course, means that there is a 15 percent chance that the app will not work, in which case the company won't be able to communicate with its delivery agents and may not be able to deliver food for at least two days. The company's daily revenue is on average $50,000.

Would you recommend to implement the app?

a) Yes, because the expected value is $240,000.

b) Yes, because a risk of 15 percent is acceptable.

c) No, because the expected value is too low.

d) No, because a risk of 15 percent is too high.

e) I need more information.

Consideration 3: Is it the biggest bang for your buck?

You may notice that the PICOC element ***Comparison*** is seldom defined, as in most organizations, no comparison is considered. For a meaningful calculation of a decision's expected value, however, you need a point of comparison. After all, in evidence-based management, the question is not so much "Does X have an effect on Y?" as it is "Does X have a larger effect on Y than Z does?" For example, in medicine, the question is not whether a new medicine has a positive effect on a certain disorder but whether it works better than existing medicines. For example, when a traditional sleeping pill increases a person's amount of sleep by 60 minutes, the added value of a new pill that extends this time by only 5 minutes is limited, especially when this new pill has side effects.

The same is true of new management methods and "cutting-edge" techniques. As explained in *Module 6, A Short Introduction to Science*, because of placebo effects, many methods and techniques have a positive effect on organizational outcomes anyway. The question is whether these new methods and techniques work better than do existing ones. For example, many companies invested lots of time and money in the now popular notion of employee engagement, as it has been widely believed that engaged workers are likely to perform better than their disengaged peers. And indeed, there are empirical studies suggesting engagement has a positive correlation with performance. However, this correlation is similar to that of employee satisfaction, a rather traditional but more valid and reliable construct. (5) In addition, other existing constructs such as social cohesion, information sharing, and goal clarity tend to have a substantially higher correlation with performance than does engagement, thus giving you more bang for your buck. (6)

Other things you can consider regarding new methods and techniques are the ease of implementation, speed, and feasibility. Option A may have a slightly bigger impact on the desired outcome than option B, but when option A requires a 2-week training of the company's workforce, option B, which does not, may be the better choice.

Learn by doing 13.3

Read this excerpt from a newspaper article

A new meta-analysis published in the **Journal for Organizational Behavior** *shows that employees high in impulsivity are more likely to engage in unethical behavior. In addition, it was found that they are also more willing to comply with supervisors' requests for compliant misconduct. This finding is consistent with self-regulation theory, which maintains that impulsive individuals have less willpower and self-control.*

Imagine you are responsible for the recruitment and selection of new employees for a bank. Would you test candidates for impulsive behavior?

a) Yes, because ethicality is very important for a bank, and according to the new study, impulsive behavior is a good predictor.

b) No, because this is a new study. New studies cannot be considered to be strong evidence because the findings can be falsified by follow-up studies; the new study is m ely indicative.

c) No, because the meta-analysis most likely reported only correlations, so the conclusion that employees high in impulsivity are more likely to engage in unethical behavior is premature.

d) I need more information.

Did I get this 13.2

Read the scenario below

A small but fast-growing international company with offices in Finland and the United States needs to hire new employees. The company has 23 employees, but the founder-CEO expects this number to increase to 40 within a year. The CEO feels that the current selection procedure (testing candidates' general mental ability [GMA, also called general intelligence] and conducting structured interviews) is not sufficient and therefore suggests screening candidates through an assessment center instead. When you consult the research literature, you find a recent meta-analyses that summarizes what 85 years of research in personnel psychology has revealed about the predictive validity of selection methods. In this meta-analysis, a table is included that lists widely used selection methods and their correlation with job performance. In addition, the table shows the added value (% gain) when a selection method is combined with a GMA test. A copy of this table is depicted below.

Based on the evidence presented in the table, would recommend replacing the company's current selection procedure (GMA test and structured interviews) with a procedure in which candidates are screened by an assessment center?

Predictor	Correlation (r)	% Gain if combined with GMA
1. GMA test	0.65	
2. Structured selection interviews	0.58	18%
3. Reference check	0.26	8%
4. Biographical data	0.35	6%
5. Years of job experience	0.16	5%
6. Person-organization fit	0.13	4%
7. Person-job fit	0.18	2%
8. Assessment center	0.36	2%
9. Years of education	0.10	1%
10. Age	0.00	0%

a) Yes, because assessment centers measure a broader range of knowledge, skills, and abilities than more traditional methods, and the selection is usually done by professional assessors with a lot of experience.

b) No, because assessment centers cost a lot of money but add hardly any value.

c) I need more information.

Consideration 4: Is the level of risk acceptable?

Calculating – or estimating – the expected value helps you to make better decisions. It helps you to decide whether the probability of a certain outcome makes taking a risk worthwhile. The probability that a negative outcome will occur represents risk. Risk acceptance, also called ***risk appetite***, is the level of risk a manager or organization is willing to accept. Individual people and organizations, however, don't value probabilities in the same way and thus have different risk appetites. Entrepreneurial people, for instance, may place high value on the small probability of a huge gain and low value on the larger probability of a loss.

In contrast, an administrator of a public organization funded with taxpayers' money may place little value on the probability of gain because his or her strategic goal may be to preserve the organization's capital. This difference stems from the difference in risk acceptance/risk appetite. Many managers, leaders, and policymakers determine in advance the level of risk they consider acceptable. A manager's level of risk acceptance, however, is affected by many factors, such as (perceived) accountability, timing, context and individual perceptions. All of these factors affect the extent to which a manager perceives a risk as acceptable. As a result, CEOs, executive boards, steering groups, and individual managers can sometimes reach very different decisions on the basis of the same evidence.

A few (more) words on probabilities

In the module *AGGREGATE: Weigh and Pull Together the Evidence*, we explained that evidence-based management is all about probabilities: the extent to which something is likely to happen or to be the case – given the evidence – expressed as a percentage. The higher the probability of an outcome, the more certain that the outcome will occur. For example, if we find that the probability of a claim (e.g., "engaged workers perform better than their disengaged peers"), given the available evidence (e.g., a meta-analysis), is 20 percent, which means there is an 80 percent probability that this claim is incorrect.

The problem with probabilities, however, is that they are hard to grasp. This is especially true of what is referred to as a single event or outcome. Most people understand that when they throw a pair of dice 100 times, they will get some sixes. But when they see the probability of a single event or outcome happening – which is the case with many claims, assumptions, or hypotheses in the realm of management – they tend to think, *Is this true? Is this going to happen – or not?* In an attempt to make sense of probabilities, most people round them to either 0 or 100 percent, which is misleading.

That's what many Americans did when they heard that Donald Trump had a probability of 10 to 20 percent of winning the 2016 presidential election. (7) When he won, many people complained that the probabilities were wrong. They were not. Just because a probability is low, it does not mean it won't happen (8). In daily life, we experience improbable events all the time: we unexpectedly meet friends in unlikely places, we win a prize in the lottery, and our new car breaks down just when we have that important job interview. As an evidence-based manager, you should resist saying that a claim, assumption, or hypothesis is "right" or "true" if its probability is above 50 percent and "wrong" or "untrue" if it is below 50 percent. A thorough understanding of probabilities and being able to calculate a decision's expected value and level of risk is a prerequisite for evidence-based decision-making.

Did I get this 13.3

Read the following scenario

A large international brewery produces beer that is sold in 12 countries across Europe. The beer is brewed in the United Kingdom and then transported to a large distribution station in Belgium where the beer gets bottled and distributed. The company has noticed that in the past year, the number breakdowns and stops in the bottling process have increased. To improve this situation, the company considers introducing Q-management, a business process improvement program widely used in many industries. The company's risk manager estimates the chance that the new program will have a negative impact on the quality of the beer to be 5 to 10 percent, which is considered acceptable. Within a week after the implementation, however, the bottling process must be stopped twice because batches of beer are contaminated – resulting in an operational loss of $2 million. The company's CEO considers firing the risk manager – obviously, his risk estimate was totally inaccurate

Do you agree that the risk manager's estimate was inaccurate?

a) Yes. The fact that within a week, the bottling process had to be stopped twice indicates that the risk estimate was too low.

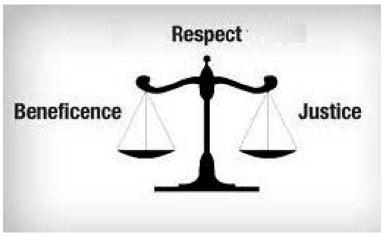
b) No. The fact that within a week, the bottling process had to be stopped twice does not indicate that the risk estimate was too low.

c) I need more information.

Consideration 5: Are there ethical issues to consider?

Decisions have important ethical implications. For example, the possibility that a decision may harm a particular group of stakeholders – a department whose already stressful workload increases, the environmental impact of a product packaging change that increases waste and pollutants, or the neighborhood facing greater noise or traffic f om a facility expansion – raises ethical concerns when we consider a decision's full ramifications. A key issue is not that a decision may have negative outcomes: organizational changes often have winners and losers and even successful decisions can involve considerable effort and difficulty for the people involved. Rather, ethical concerns arise when a decision's benefits and harms aren't evenly distributed across stakeholders. Cutting jobs and employee pay while increasing compensation to senior executives may raise issues regarding perceived justice (a key issue in ethics) as well as possible harms to the organization in terms of reputation and other market-related effects. When considering a decision's ethicality, three ethical principles can help managers reflect on the implications of the decisions they make: beneficence, respect, and justice.

Beneficence



Beneficence deals with the welfare of stakeholders, particularly whether the benefits from a decision exceed the harms that it might bring. Certain organizational decisions involve legitimate harms, such as putting a competitor out of business, raising prices, or closing a line of business. We often regard such outcomes as the price of doing business. However, when the organization's mission or corporate values have committed it to quality relationships with its suppliers and customers, such harms may be perceived as violating the principle of beneficence, especially when they are deemed avoidable. Decisions by drug companies to raise prices without a compelling economic need to do so can fall into this category. On the other hand, where the ecological impact of a decision is concerned, Greenpeace might be expected to make different decisions than Shell makes as a result of their distinct corporate values, mission, and stakeholder interests.

Respect for persons

Respect for persons is the ethical principle promoting the exercise of autonomy in managing one's life and making personal decisions. It involves avoiding imposing undue demands or risks that undermine an individual's well-being and providing appropriate information to help people make choices that reflect their own goals and interests. For example, managers who fail to inform stakeholders about the potential consequences of working with hazardous substances such as asbestos would violate the principle of respect for persons.

Justice

Justice is the ethical principle to treat people equitably and distribute benefits and burdens fairly. One of the major concerns in organizational decisions is when decisionmakers ignore adverse effects on parties "not in the room" or stakeholders with little power and voice. Disproportionately benefiting one set of stakeholders while another is harmed raises issues of injustice. For example, Uber – a global taxi company in the United States – was stripped of its London license because the organization disregarded the values and concerns of important stakeholders (employees, competitors, local community) and ignored regulations protecting the interests of conventional taxi companies. (9)

Learn by doing 13.4

As a decision maker, you will most likely be faced with situations in which it is unclear what is the right or ethical thing to do, particularly when doing right is different from doing what's most profitable. Of course, when a decision will seriously hurt a specific group of stakeholders, the choice to make is often clear. But many decisions fall into the gray area in between. Read the two examples that follow and answer the questions they pose. Obviously, there is no right or wrong here, but when considering a decision's ethicality, think about the three ethical principles discussed above: beneficence, respect, and justice.

By introducing a new production technique, the manufacturing costs of a medical device your company makes has dropped by 50 percent. The sales revenue of the medical device is your company's only source of income, and because of the reduction in production costs, the company will finally make a healthy profit. One of the company's customers, a large hospital chain, asks you for a discount. The hospital's head of purchasing tells you he knows about the 50 percent cost reduction because he is best friends with your company's vice president. Your director of sales approves the discount. A few days later, two other customers, a small local community hospital and an academic hospital, place the same order.

Do you tell the other customers about the drop in manufacturing costs, and do you offer them a similar discount?

a) Yes

b) No

Learn by doing 13.5

Read the following scenario

You are asked by an important client, a multinational firm with a large budget for consulting services, whether your consulting firm can deliver a two-day MBTI workshop for its top 100 executives. Your initial enthusiasm about this 100k + assignment wears off, however, when you look into the scientific literature on MBTI. You learn that MBTI stands for Myers Briggs Type Indicator, a popular personality test that is seriously flawed and gives an oversimplified and non-evidence-based view of human personality. In contrast, the Big 5 Inventory is the most valid and reliable model of personality, supported by a large body of scientific evidence and widely used by psychologists.

Most likely, you will explain to your client that the MBTI is seriously flawed and recommend the use of the Big 5 instead. But what would you do if the client still insists on using the MBTI and threatens to go to another consulting firm? Would you accept the assignment and do the workshop?

Learn by doing 13.6

Read the scenario we used earlier

Introducing a new production technique, the manufacturing costs of a medical device your company makes has dropped by 50 percent. One of the company's customers, a large hospital chain, asks you for a discount; the hospital's head of purchasing tells you he knows about the 50 percent cost reduction because he is best friends with your company's vice president. Your director of sales approves the discount. A few days later, two other customers, a small local community hospital and an academic hospital, place the same order. You decide not to tell the other customers about the drop in manufacturing costs and not to offer them a similar discount.

Which ethical principle is of concern here?

a) Beneficence
b) Respect
c) Justice

Consideration 6: Is the evidence actionable?

Example

Imagine you are an executive at a large Italian hotel chain. You are thinking about reducing the number of middle managers and granting local teams more autonomy. Before you implement this change, you want to make sure that collaboration within the company's teams is at its best, so you ask the question, ***Which factors positively affect team collaboration?*** *When consulting the scientific literature, you find that one of the most important antecedents of team collaboration is trust among team members.* (10) *The company's experienced managers and team members confirm this finding and assert that interpersonal trust is indeed a prerequisite for a well-functioning team. Since the scientific evidence stems from two meta-analyses representing 80 samples from various industries, you judge the evidence to be applicable to your organizational context.*

To measure the teams' level of interpersonal trust, you apply a short questionnaire that researchers widely use. The outcome indicates that some teams score well below average. You therefore decide to try to improve this situation. However, you now face a new question: How can you increase trust among team members? The scientific evidence explains what team trust is and how you can measure it, but it fails to explain how team trust can be increased.

In this example, the evidence you have in hand is actionable only for the diagnostic stage of the decision-making process; that is, Do we have a problem with trust within our teams, and how can we measure this? For the solution stage – How can we increase trust among team members – its applicability is rather limited. You can't act on it. As a result, you will have to consult the scientific literature for a second time and search for studies on factors that increase intra-team trust.

The problem of nonactionable evidence often occurs when only cross-sectional studies are available, in which only a correlation between two variables is measured. In contrast, controlled longitudinal studies in which the independent variable (e.g., trust) is manipulated by the researchers often provide more actionable information. The same is true of case studies describing organizational interventions and practices. Although the internal validity of this type of study is rather low, such studies can provide practical examples of how findings might be applied.

TIP...

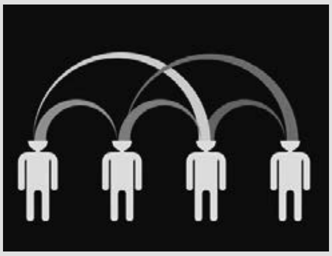
Implications for practice

When dealing with evidence from research, make sure you consult the section in research articles called *Implications for Practice*. Most research articles include it, and in general, such sections are informative for management practice. However, be critical if general terms such as *focus*, *align*, or *enhance* are used. These terms suggest actionability, but the recommendation "Our findings suggest that in o der to increase cooperation within a team, managers should focus on enhancing intra-team trust" is correct from an academic point of view but useless from a practical perspective. You should always critically assess whether the constructs referred to in the Results section or Implications for Practice are actionable and represent interventions you can use in practice.

When the evidence from the scientific literature is not actionable, evidence from practitioners and stakeholders can help. Always ask employees and managers for their professional judgment and experience. Experienced practitioners often have a wealth of practical knowledge, and although this knowledge may be based on individual experience, it may provide indispensable insights when it comes to practical application of research findings. Relevant questions to ask a e

- How can we apply this?
- What should we take into account?
- What do we need in order to apply this?

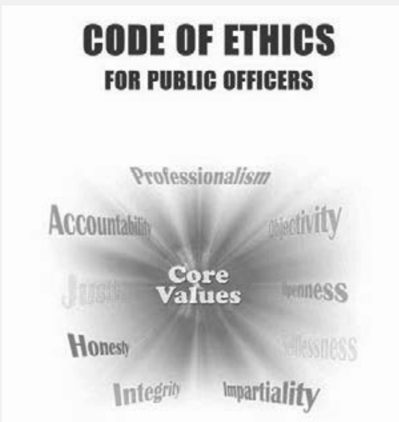
Example



*An international IT firm wants to improve the performance and timeliness of its software development teams. When consulting the scientific literature, the firm finds that one of the most important antecedents of team performance is information sharing. Several studies indicate that, especially if complex problems have to be addressed, information sharing allows team members to pool their knowledge and past experiences, which is particularly important for the generation of new ideas (11). The firm's managers and software developers confirm that information sharing between team members is indeed important and acknowledge that this may be an issue within some of the company's teams. To check this assertion, the teams' level of information sharing is measured with a short questionnaire that researchers often use. The outcome confirms that several teams score well below average. It is therefore decided that action should be taken to improve this situation. The available evidence, however, fails to provide guidance for possible actions. For this reason, the firm's senior managers are consulted by means of a Delphi procedure (see the module "*ACQUIRE – Evidence from Practitioners*"). The outcome indicates that a variety of methods are available to enhance information sharing, such as show-and-tell sessions or informal gatherings where members can share their expertise. Another relatively new but promising method is a daily scrum: a short stand-up meeting at the beginning of the day during which team members share information and speak up about any problems that might prevent project completion. Based on this information, the company decides to inform all teams about the value of information sharing and the various ways in which it can be increased and to ask the teams to choose for themselves a method that they consider the most effective and feasible.*

Learn by doing 13.7

Read the following four findings from a systematic review of the research literature on predictors of unethical behavior in the workplace, and determine whether each finding is actionable.



1. *A randomized controlled study found that having a range of different tasks to do in one's job is negatively associated with unethical behavior. In this study, participants were assigned to a low task variety or high task variety condition, and then presented with a scenario in which they had the opportunity to cheat. The authors found that those with higher task variety were less likely to cheat than those with low task variety. These findings suggest that feelings of disinterest toward one's job can translate into unethical behavior.*

a) Actionable

b) Not actionable

2. *A recent controlled study indicated that, when the organization is in competition with other organizations, individuals are more inclined to engage in pro-organizational unethical behavior when faced with ethical dilemmas.*

a) Actionable

b) Not actionable

3. *Two randomized controlled studies demonstrated that when supervisors are held accountable for their actions, they are more likely to disapprove of unethical behavior by their subordinates.*

a) Actionable

b) Not actionable

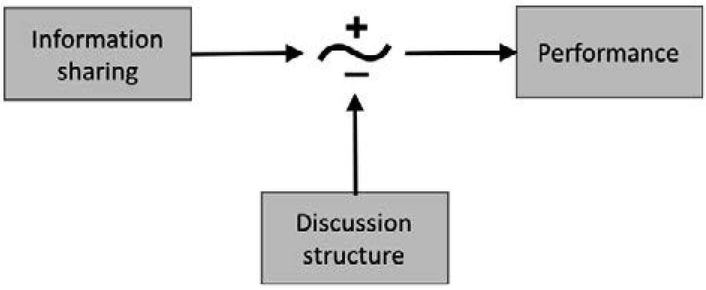
4. *A code of ethics or code of conduct provides guidelines and expectations regarding ethical behavior within organizations. However, a code will not guarantee ethical behavior and choices. In a meta-analysis of 136 studies, it was shown that the mere existence of a code of ethics does very little to discourage unethical behavior. Rather, it was found that it is the enforcement of a code of ethics that has a moderate effect on decreasing unethical behavior.*

a) Actionable

b) Not actionable

Consideration 7: Are there moderators to take into account?

Another important aspect that you need to consider when applying evidence is the effect of moderators. As explained in the module *ACQUIRE Evidence from the Scientific Literature*, a moderator is a variable that affects the direction and/or strength of the relation between a predictor (e.g., information sharing) and an outcome (e.g., performance). Put differently, moderators indicate when or under what conditions we can expect a particular effect. For instance, in the example we used earlier, the evidence suggests that information sharing has a larger (positive) impact on performance when discussions within the team are structured and focused. (11) Thus, the positive effect of information sharing on performance is moderated by the factor "discussion structure."



Often, there are several moderators that affect an outcome. As a result, moderators are important success factors for the application of the evidence when addressing a problem or implementing a solution. You need to take into account moderators from the organizational context, such as industry sector, team size, and task type, because they may weaken or strengthen the effect of an intervention or practice. For example, a healthcare organization facing an issue of low productivity may find evidence that financial incentives increase motivation and productivity. However, when most of the organization's staff works in groups, the evidence needs careful examination, as the effect may be moderated by whether the incentive is based on individual or team performance. (12)

Did I get this 13.4

Read the following three findings from a systematic review of the research literature on the effectiveness of diversity training and determine whether there is a moderator that needs to be taken into account.

1. *A meta-analysis based on 236 studies demonstrated that diversity training in an educational setting elicits stronger effects than such training does in a corporate environment. A possible explanation may be that employees in organizations see diversity training as an "add-on" practice, something that "takes time away from work" and is secondary to the purpose of the organization. In contrast, diversity training in an educational setting is usually part of the mission of such institutions, so it may be perceived as an opportunity to learn about diversity and prejudice and apply concepts through experiential learning.*

a) Yes, there is a moderator.
b) No, there is no moderator.

2. *A recent meta-analysis indicates that diversity training elicits intense emotional responses, and participants see the training as effective and worthwhile. Reactions to training are an antecedent of learning that in turn leads to the desired behavior in the short term. However, some people might like the training for reasons that are not directly related to its content (e.g., the trainer's sense of humor), and as a result, their behavior will not change.*

   a) Yes, there is a moderator.
   b) No, there is no moderator.

3. *A recent meta-analysis found that trainees' motivation is higher and effects on attitudes and behaviors are stronger when a direct manager/supervisor delivers the training, as opposed to an external trainer, a diversity and inclusion manager, or an HR generalist.*

   a) Yes, there is a moderator.
   b) No, there is a moderator.

How and in what form can you apply the evidence?

Managers and leaders make a wide variety of decisions. Some decisions are rather mundane, such as sending an email, booking a room, or drawing up an agenda. Other decisions may have a huge impact on the organization, such as the decision to initiate a hostile takeover of a large competitor. As a rule, the type of decision determines how you can effectively apply the evidence in the decision-making process.
Management decisions can be classified in many ways – the most common is probably the distinction among strategic, tactical, and operational decisions. The classification below is not completely clear cut, but it provides you a guide in effectively applying evidence in decision-making.

Routine decisions

Often operational, routine decisions are those made using an organization's existing set of rules and procedures. They may not have a major impact on resource allocation and are typically made by middle or first-line managers: How much of a given product to stock in a shop? Who to hire to fill an existing ole? How to deal with a customer complaint? Routine decisions are routine because we make them repeatedly. They concern the typical problems every manager faces on a daily or weekly basis. Often, they involve performance feedback, goal setting, conflict management, motivating employees, recruitment and selection, sales performance, absenteeism, and so on – all issues for which much research evidence is available. Important reasons for taking an evidence-based approach to making routine decisions are to get good results more consistently, to find ways of improving results, and to free up time to make other decisions better.

We typically apply evidence for routine decisions by using a "push" approach: actively distributing information to the organization's relevant stakeholders. Such evidence may take the form of procedures known to be effective, where repeated practice demonstrates what works and what does not. This evidence can be provided in easy, accessible, and user-friendly forms such as a protocol, checklist, flowchart, decision tree, or standard operating procedure (SOP).

Checklists and protocols are widely used in high-risk industries, such as aviation and healthcare. For example, to reduce the number of medical errors in surgery, the World Health Organization's "Safe Surgery Checklist" was introduced in 2008. This checklist comprises a simple time-out procedure before the start of the surgery, where the surgical team checks the patient's name, the intended medical procedure, and the patient's site (e.g., left or right kidney). In 2009, the WHO concluded that the use of the checklist is associated with a significant dec ease in postoperative complications (30%) and mortality rates (50%). (13)

TIP...

Avoid over-simplification and over-standardization!

When applying evidence in the form of a checklist, protocol, or SOP, it is important to avoid over-simplification and over-standardization. (14) Although technical rationality may be needed in a cockpit or operating room, it can be too rigid for informing managers about how to give performance feedback to employees or how to select candidates in an unbiased way. Keep in mind that all checklists, protocols, and SOPs are incomplete. They can help to guide a decision or practice, but users need to remain mindful of the specific context. (15) The goal is to create a guideline that is agile where it needs to be. It is important that people understand the principles behind the guideline and know how best to go off script if the situation warrants. Application always depends on the right balance between standardization and informed judgment.

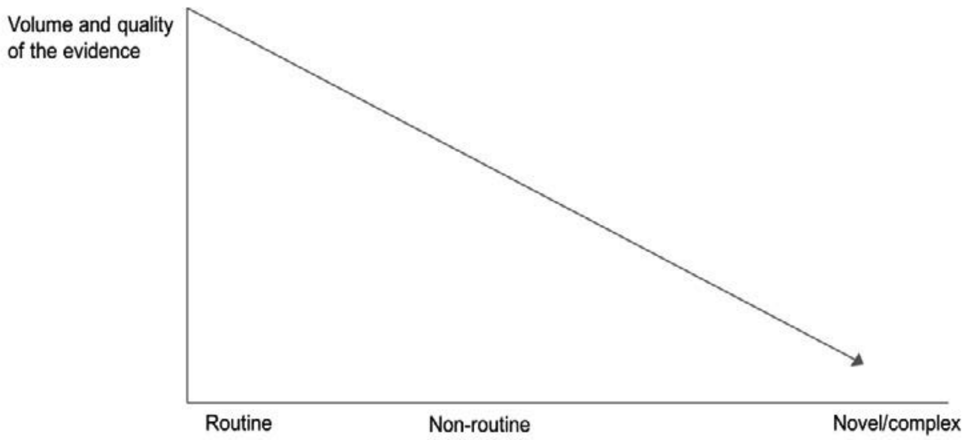
Nonroutine decisions

Nonroutine decisions involve making changes to the way in which an organization operates. They are the sort of decisions that managers are typically trained to make at business schools and universities. The scale of this type of decision can vary from the introduction of autonomous teams within a particular department to major strategic decisions such as whether to take over another company or launch a new product. The key point about decisions that fall within this category is that they concern something that is new to the organization – though not necessarily new to the industry or to other firms

Consequently, they are subject to varying degrees of uncertainty regarding the outcome. Nonroutine also means that relevant research evidence may be available, but it may take effort to retrieve it. In nonroutine decisions, the evidence needed must first be identified, and therefore, the first action often concerns understanding the problem (or opportunity) and defining the desired outcome.

In contrast to routine decisions, evidence for nonroutine decisions is typically applied using a "pull" approach: based on the problem to solve, the outcome to achieve, and the organizational context involved, you can actively obtain evidence from multiple sources and succinctly summarize it to inform decision-makers. Evidence summaries come in many forms. When it comes to summarizing scientific evidence, critically appraised topics (CATs) and rapid evidence assessments (REAs) are the forms we most widely use. Both apply the same systematic approach to selecting the studies: the methodological quality and practical relevance of the studies are assessed on the basis of explicit criteria. Thus, the summaries are transparent, verifiable, and reproducible. CATs are the quickest method and may take one skilled person a few days to produce. (16) REAs might take two skilled persons several days or weeks to produce.

Although there is no harm in drawing up a separate summary for each evidence source (practitioners, organizational data, stakeholders, and the scientific literature), we recommend aggregating and synthesizing the evidence into one overall summary.



Novel and/or hypercomplex decisions

These decisions involve interventions that are new not only to the organization but also to the industry at large (e.g., introducing an innovative IT solution or starting a business unit in an emerging market). They often involve new or emergent conditions for which prior experience and historical knowledge provide little insight. Imagine a major technological breakthrough or a heretofore unheard of environmental catastrophe. In such circumstances, evidence from the scientific literature is often not available because the issues are too novel for a scientific study.

As a result, this type of intervention is typically subject to a high degree of uncertainty and involves many unknowns. Novel or hypercomplex situations involve considerable ambiguity – cues that signal the problem or possible solution can be so vague and confusing that they are hard to recognize. Such can be the case when technological changes come with unexpected consequences, including new opportunities or unimaginable threats. Often in these situations, little or no quality evidence is available, so there is insufficient evidence to inform decision-makers and no choice but to work with the limited evidence at hand and supplement it through a process of sense-making and learning by doing. This means pilot testing and systematically assessing the outcomes of our decisions. In highly uncertain organizational environments, managers and leaders are likely to rely on constant experimentation and critical reflection in o der to identify which things work and which things do not. We discuss how you can do this in the module *ASSESS: Evaluate the Outcome of the Decision*.

Learn by doing 13.8

Given the type of decision at hand, determine the best form in which the evidence can be applied.

1. *An insurance company that wants to increase the sales performance of its sales agents.*

a) A protocol, decision tree, or standard operating procedure
b) An evidence summary, such as critically appraised topic (CAT) or rapid evidence assessment (REA)
c) Probably little or no evidence is available, so this requires a process of learning by doing or pilot testing and systematically assessing the outcome

2. *An insurance company that wants to develop a new line of products related to underserved and emerging markets.*

a) A protocol, decision tree, or standard operating procedure
b) An evidence summary, such as critically appraised topic (CAT) or rapid evidence assessment (REA)
c) Probably little or no evidence is available, so this requires a process of learning by doing or pilot testing and systematically assessing the outcome

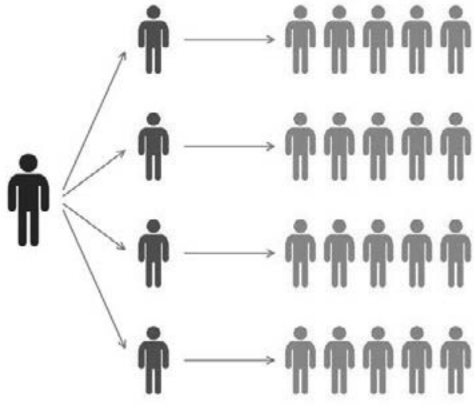
3. *An insurance company that wants to improve its recruitment and selection procedure of sales agents.*

a) A protocol, decision tree, or standard operating procedure
b) An evidence summary, such as critically appraised topic (CAT) or rapid evidence assessment (REA)
c) Probably little or no evidence is available, so this requires a process of learning by doing or pilot testing and systematically assessing the outcome

4. *An insurance company that wants to reduce the absenteeism of its sales agents.*

a) A protocol, decision tree, or standard operating procedure
b) An evidence summary, such as critically appraised topic (CAT) or rapid evidence assessment (REA)
c) Probably little or no evidence is available, so this requires a process of learning by doing or pilot testing and systematically assessing the outcome

Final thoughts: Dissemination, implementation, and change



The methods of applying evidence mentioned on the preceding page all represent different ways to provide decision-makers with the best available evidence from multiple sources. In most evidence-based disciplines, however, dissemination of evidence is a major concern. For example, a survey of 950 American HR managers showed large discrepancies between what managers think is effective and what the current scientific research shows. (17)

In medicine, it has been widely reported that evidence takes on average 10 to 15 years to be incorporated into routine healthcare practice. The situation in other evidence-based disciplines, such as education, policing, and social welfare, is not much better. (18) Whether the situation relates to a guideline, a checklist, or a CAT, simply providing practitioners with the best available evidence unfortunately does not guarantee that they will use it.

In fact, practitioners – deliberately or unwittingly – disregard evidence for many reasons. As a result, a new field of science has emerged: implementation science, the study of methods to promote the uptake of evidence in routine practice. Whereas dissemination concerns the spread of evidence about a model, topic, or intervention within an organization or a profession, implementation science concerns the uptake and use of that evidence. (19)

We must point out the distinction between implementation of evidence and implementation of a management model, method, or technique such as lean management or a new IT system. The latter concerns methods or insights to facilitate organizational change and is thus related to the field of change management. Obviously, there is considerable overlap between the two fields, but whereas the implementation of evidence and evidence-based management is within the scope of this module, implementation as organizational change is not. Nevertheless, we would like to dedicate a few words to the latter. (20)

Organizational change is risky. At least, that is what many change experts, consulting firms, and management gurus claim. (21) For most experts, the reason is obvious: change initiatives fail because there is no sense of urgency, because there is no clear vision, because there is no commitment to the change goals, because the change leaders lack emotional intelligence, because . . ., because . . ., and so on. Just as in fields such as marketing, leadership, and human resources, many (mostly self-proclaimed) change experts, without providing quality evidence, make strong claims about what does and doesn't work.

Here too is a huge gap between what change experts think is effective and what the evidence shows. The only response to this dilemma is to take an evidence-based approach: conscientiously acquiring and appraising the best available evidence to determine whether claims regarding the best way to change or implement stand up to evidence-based scrutiny.

To sum up...

We started this module by pointing out that, in most cases, the answer as to whether to apply the evidence and act upon its findings is not a simple yes or no. Usually, the answer is, “It depends.“ In general, seven questions need to be answered to determine if (and how) the evidence can be applied:

1. ***Is the evidence generalizable to your organizational context?***

To determine whether this context may be different from the context in which the evidence was acquired, we re-introduced the mnemonic we discussed in previous modules: PICOC. The underlying thought of PICOC is that all five elements (Population, Intervention, Comparison, Outcome, and Context) should be considered when determining whether the evidence is sufficiently generalizable to your organizational context.

2. ***What is the expected value?***

The decision whether to implement a solution depends on the notion of expected value: the (often monetary) outcome expected from the decision. The expected value can be calculated with the following formula P(benefits) + P(costs)

3. ***Is it the biggest bang for your buck?***

To determine whether to implement a solution, the question is not so much, *Does X have an effect on Y?* as it is, *Does X have a larger effect on Y than Z does?* For example, a new promising management method may have a positive effect on a certain organizational outcome, but the question is whether this effect is larger than that of existing methods.

4. ***Is the level of risk acceptable?***

Risk acceptance, also referred to as *risk appetite*, is the level of risk a manager or organization is willing to accept. Individuals and organizations have different risk appetites that are influenced by many factors, such as (perceived) accountability, timing, context, and individual perception. As a result, people can sometimes reach very different decisions on the basis of the same evidence.

5. ***Are there ethical issues to consider?***

When considering a decision’s ethicality, three ethical principles may help you reflect on the implications of a decision: *beneficence* (whether the benefits f om a decision exceed the harms it might bring), *respect* (the ethical principle promoting the exercise of autonomy in managing one’s life and making personal decisions), and justice (the ethical principle to treat people equitably and distribute benefits and burdens fairly).

6. ***Is the evidence actionable?***

Unfortunately, not all evidence is actionable. This is a particular concern with scientific evidence. It may demonstrate that A has an effect on B, but it often fails to explain how A can be increased to impact B. In those cases, the Implications for Practice section in research articles may help. Another option is to consult experienced professionals. Relevant questions to ask are, How can we apply this? What should we take into account?, and What do we need to apply this?

7. Are there moderators to take into account?

A moderator is a variable that affects the direction and/or strength of the relationship between a predictor and an outcome. Often, there are several moderators that affect an outcome. As a result, moderators are important success factors for the application of the evidence.

Finally, we discussed how and in what form you might apply the evidence.

When the decision to make concerns an operational or routine decision, we typically apply evidence using a *push* approach: actively distributing information to the organization's relevant stakeholders in the form of a checklist, protocol, or standard operating procedure.

When the decision involves making changes to the way in which an organization operates (nonroutine decisions), evidence is typically applied using a *pull* approach: actively obtaining evidence from multiple sources and succinctly summarizing it in the form of a CAT or REA.

Finally, for decisions that involve interventions that are not only new to the organization but also new to the industry at large (novel and/or complex decisions), evidence is often not (yet) available. In these cases, there is no other option but to work with the limited evidence at hand and supplement it through a process of sense-making and learning by doing.

Podcast

In this podcast host Karen Plum discusses with Eric Barends, Managing Director, Center for Evidence-Based Management , Denise Rousseau, H J Heinz University Professor, Carnegie Mellon University, Jonny Gifford, Senior Researcher, CIPD, David Creelman, CEO of Creelman Research, and Stefanie Nickel, Global Head of Diversity & Inclusion, pharmaceutical manufacturer Sandoz, how to incorporate evidence into the decision making process.

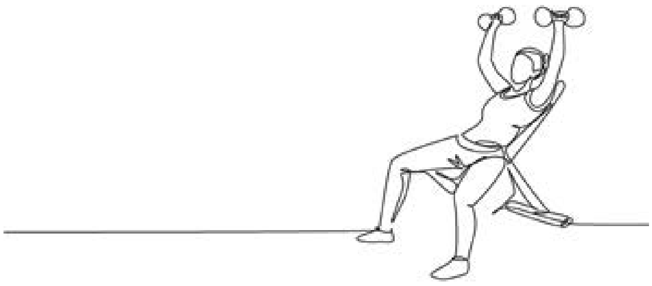
Is the evidence appropriate for our situation (is it generalizable)? Can we (and should we) action the evidence? And what are the practical aspects associated with using the evidence that we've gathered? The real world is messy and complex, so there are many things to take into consideration.

We share real world examples from professional HR body CIPD, Creelman Research and a large trial at pharmaceutical organization Sandoz (part of Novartis Group), which shows not only the power of capturing stakeholder experiences, but the value of keeping interventions simple and practical.

Karen Plum   Jonny Gifford   Denise Rousseau   Eric Barends   David Creelman   Stephanie Nickel

https://evidencebasedmanagement.buzzsprout.com

Exercise Module 13



Exercise 13.1: Performance feedback – applying the evidence

Many organisations expect their managers and supervisors to provide their subordinates with feedback: information about their performance that serves as a basis for recognition, encouragement, and/or improvement. Indeed, a review of a large number of scientific studies indicates that feedback can have a positive effect on employees' learning and performance.

Download an ***abbreviated version of the review here*** and write down your answers to the following questions:

1. Is the evidence presented in the review generalizable to your organizational context (PICOC)? Why/why not?
2. Is the evidence, in particular the seven findings, actionable
3. Are there contextual factors that you feel need to be taken into account?
4. How and in what form would you apply the findings? Think about guidelines, checklists, presentations, workshops, trainings, or other forms that you could use to disseminate the evidence within your organisation. Write a proposal (for your boss, HR, or other audience) in which you describe in detail the design, content, and target audience.

Submit a Word document with your answers (max. two pages) according to the instructions of your professor/instructor.

Suggestions for further reading

- *Protocol-based care: the standardisation of decision-making?* Rycroft-Malone et al, JCN, 2009
- *The Importance & Value of the CHECK LIST*. Atul Gawande, TED talk, 2015
- *An introduction to implementation science for the non-specialist*. Bauer et al, BMC Psychology, 2015
- *GRADE: going from evidence to recommendations*. Guyatt et al, BMJ, 2008
- *Rapid Evidence Assessments in Management – Example & application*. Barends et al, 2017
- *Evidence-based change management*. Rousseau, D. M., & ten Have, S., Org Dynamics, 2022

References

1. CIPD (2016). *Could do better? Assessing what works in performance management* [research report]. https://www.cipd.co.uk/knowledge/fundamentals/people/performance/what-works-in-performance-management-report.

2. In science, generalizability is also referred to as external validity. Often, two types of external validity are assessed: ecological validity (type of organization) and population validity (type of employees).

3. Drosdeck, J., Bansal, S., Peterson, L. A., Arnold, M., Harzman, A., Muscarella, P., & Ellison, E. C. (2013). A personality trait analysis of surgeons and nonsurgeons at different career points. International Conference on Residency Education Conference Research Abstracts. *Royal College of Physicians and Surgeons of Canada and Journal of Graduate Medical Education.*

4. Closely related to the notion of expected value is the concept of expected utility. Whereas expected value simply calculates the monetary value, expected utility also takes into account the personal satisfaction and happiness that is derived from a certain outcome.

5. Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). "Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis." *Journal of Applied Psychology*. Volume 87. Number 2. 268-279 Pages.

6. Barends, E., Plum, K., & Mason, A. (2017). Rapid evidence assessments in management: An example. In A. R. Kovner & E. A. D'Aunno (Eds.), *Evidence-based management in healthcare*. Health Administration Press. Chicago. Pages.

7. Katz, J. (November 8, 2016). "Hillary Clinton has an 85% chance to win." *New York Times*.

8. Leonard, D. (December 24, 2017). "What I was wrong about this year." *New York Times*.

9. Butler, S. & Topham, G. (September 22, 2017). "Uber stripped of London license due to lack of corporate responsibility." *The Guardian*.

10. McAllister, D. J. (1995). "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations." *Academy of Management Journal*. Volume 38. Number 1. 24-59.

11. Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). "Information sharing and team performance: A meta-analysis." *Journal of Applied Psychology*. Volume 94. Number 2. 535-546.

12. Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L., Furler, J., & Young, D. (2011). "The effect of financial incentives on the quality of health ca e provided by primary care physicians." *Cochrane Database of Systematic Reviews*. Volume 9. September 7 Pages. CD008451.

13. World Health Organization (2011). "*New scientific evidence supports WHO findings: Surgical safety checklist could save hundreds of thousands of lives*." www.who.int.

14. Rousseau, D. M. (2018). Making evidence-based decision in an uncertain world. *Organizational Dynamics*. https://doi.org/10.1016/j.orgdyn.2018.05.001.

15. Rycroft-Malone, J., Fontenla, M., Seers, K., & Bick, D. (2009). Protocol-based care: The standardisation of decision-making?. *Journal of Clinical Nursing*. Volume 18. Number 10. 1490-1500.

16. A guideline on how to conduct a CAT can be found on the CEBMa website under resources and tools: https://www.cebma.org/resources-and-tools/ .

17. Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*. Volume 41. Number 2. 149-174.

18. Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding the time lags in translational research. *Journal of the Royal Society of Medicine*. Volume 104. Number 12. 510-520.

19. Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*. Volume 3. Number 3. 1-12.

20. ten Have, S., Huijsmans, A. B., & Otto, M. (2016). *Reconsidering change management: Applying evidence-based insights in change management practice.* Routledge. New York.

21. Beer, M., & Nohria, N. (2000). Cracking the code of change. *Harvard Business Review*. Volume 78. Number 3. 133-141.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 13.1

a,b: Correct

c: Defensible

d: Not quite right

e: Incorrect

Obviously, the PICOC is different here: the population (doctors versus senior managers) and the context (healthcare organizations versus financial services organizations) a e not the same. Nevertheless, we would argue that the finding that ole models have a (positive or negative) effect on the values, attitudes, and behavior of other people seems to reflect a general psychosocial mechanism applicable to all people. In addition, the finding also aligns with the social theory of reference groups, which states that individuals tend to use other people as a standard for evaluating their own behavior.

Learn by doing 13.2

When a study is conducted in an artificial setting, the kneejerk eaction is to condemn the study for questionable generalizability. Laboratory studies, however, can have great applied value because the purpose of most laboratory research is not to examine a specific population in a particular setting but to test a model or theory that is applicable across populations and settings. Put differently, it is meant as a “proof of principle.” In addition, generalizing from one setting to another, or from one population to another, is as problematic for research in a natural setting as it is for research in an artificial setting. As a result, there are natural studies with low generalizability and laboratory studies with high generalizability. Finally, several meta-analyses have found no differences between findings f om laboratory studies and natural studies on a wide range of topics relevant to management, such as goal setting, performance feedback, unethical behavior, information sharing, and leadership style. Thus, the answer to the question whether studies conducted in an artificial setting a e generalizable to the real world is: *it depends*. As such, it takes a process of “thinking through” to determine if a laboratory study (or any other study) is applicable to your organizational context.

Did I get this? 13.1

a: Correct > The expected value is 255,000 (0.85 × 300,000) minus 15,000 (0.15 × 100.000) = $240,000. Of course, you should also take into account potential intangible costs, such as a decrease in customer satisfaction or reputational damage.

b: Incorrect > Whether or not a risk is acceptable depends on many factors, such as the expected value. In this case the expected value is 255,000 (0.85 × 300,000) minus 15,000 (0.15 × 100.000) = $240,000. Of course, you should also take into account potential intangible costs, such as a decrease in customer satisfaction or reputational damage, but in this case, the expected value suggests implementing the new app is a good business decision.

c: Incorrect > see A

d: Incorrect > see B

e: Incorrect > More information is often useful, such as information about potential intangible costs, such as a decrease in customer satisfaction or reputational damage, but in this case, the expected value ($240,000) suggests implementing the new app is a good business decision.

Learn by doing 13.3

a: Incorrect > You are correct that ethicality is important for a bank, but the article fails to provide relevant information (see D)

b: Incorrect > You are correct that the findings of a new study can be falsified by follow-up studies, bu given that the study is a meta-analysis, it is less likely.

c: Incorrect > You are correct that the meta-analysis may have reported only correlations, but that information is not provided.

d: Correct > The article fails to provide information about how much more likely individuals high in impulsivity are than individuals low in impulsivity to engage in unethical behavior: 2 times more likely, 10 times, or maybe even 100 times? In addition, we do not know whether there may be other (more effective, or easier) methods to prevent people from engaging in unethical behavior. We need more information.

Did I get this? 13.2

a,c: Incorrect

b: Correct

As you can see in the table, the predictive validity of an assessment center is r = 0.36, which is lower than a GMA test (0.65) and structured interviews (0.58). Even combining a GMA test with an assessment center leads to only a very small increase of the predictive validity (2%). This means that the current selection procedure gives you the biggest bang for your buck.

Did I get this? 13.3

a: Incorrect

b: Defensible

c: Correct

The fact that within a week, the bottling process had to be stopped twice does not mean that the risk estimate of 5 to 10 percent was too low. Even a risk estimate of only 1 percent does not guarantee a trouble-free implementation. To determine whether the risk manager’s estimate was accurate, more information is needed.

Learn by doing 13.4

Yes: Correct

No: Incorrect

Your firm o fered a cost reduction to the large hospital chain because of private information its purchasing manager had access to. That decision was not made for legitimate business reasons. The decision to charge other hospitals more is at odds with the principle of justice. If that discount were allowed to stand, on what basis can it be legitimate to charge other hospitals more?

Learn by doing 13.5

Yes: Use of the MBTI probably does little harm in itself, unless the firm elies on it at the expense of more valid and reliable assessments. But consider how you might educate your client to make better-quality decisions about its training activities. The workshop could be done in an ethical way if it is part of a larger program to help improve the organization through evidence-based practices, a way of acting on the principles of both beneficence (the fir s well-being) and respect for person (giving people valid information with which to make decisions).

No: You are taking a principled stand that reflects both beneficence (the fi s well-being) and respect for persons (giving people valid information with which to make decisions).

Learn by doing 13.6

a: Incorrect > The ethical principle of beneficence involves the welfa e of stakeholders, particularly whether the benefits f om a decision exceed the harms that it might bring. This is not a concern in this scenario.

b: Incorrect > The ethical principle of risk involves avoiding imposing undue demands or risks that undermine an individual's well-being. This is not a concern in this scenario.

c: Correct > The ethical principle of justice involves treating people equitably and distributing benefits and burdens fairly. When some clients profit f om the reduced costs and others do not, justice is indeed a concern.

Learn by doing 13.7

1a: Correct

1b: Incorrect

Although the findings do not state *how* you could increase task variety, as a manager, you can do several things, such as

- Ask people whether they enjoy variety and challenge in their daily work
- Discuss with those who perceive their daily work as boring how task variety could be increased
- Promote job crafting – allowing people to alter their jobs in such a way as to better suit their skills and interests thereby increasing job variety

2a: Incorrect

2b: Correct

The finding that a competitive market may lead to an inc ease in unethical behavior is not really actionable. There may be a way to cushion this risk, but this information is not provided.

3a: Correct

3b: Incorrect

Although the finding does not state how supervisors can be held accountable, as a manage , you most likely have several options to increase your employees' accountability.

4a: Defensible

4b: Defensible

The finding indicates that me ely having a code of ethics does not suffice – it is e fective only when it is enforced. Although this finding is actionable, a clear example of how to enfo ce a code of ethics would help.

Did I get this? 13.4

1a: Correct

1b: Incorrect

According to this finding, the setting of the training is a moderator that needs to be taken into account.

2a: Incorrect

2b: Correct

In this finding, no moderator is mentioned

3a: Correct

3b: Incorrect

According to this finding, the e fect of the diversity training is moderated by the functional role of the person who delivers the training.

Learn by doing 13.8

1a,c: Incorrect

1b: Correct

Increasing the sales performance of sales agents is a nonroutine decision, so a summary of the best available evidence (e.g., from the research literature or other organizations) would be a good form to apply and disseminate the evidence.

2a,b: Incorrect

2c: Correct

The evidence on this topic is very limited and probably low in quality, so this decision involves many unknowns. It should therefore be regarded as hypercomplex. Therefore, the best approach is a process of learning by doing or pilot testing and systematically assessing the outcome.

3a: Correct

3b,c: Incorrect

Every organization deals with the recruitment and selection of personnel on a routine basis. This means that a protocol, decision tree, or standard operating procedure is the best form to apply and disseminate the evidence.

4a: Correct

4b,c: Incorrect

Every organization deals on a regular basis with the need to reduce absenteeism. It is an issue for which much evidence is available. This means that a protocol, decision tree, or standard operating procedure is the best form to apply and disseminate the evidence.

CEBMa center for Evidence-Based Management

Open Learning Initiative

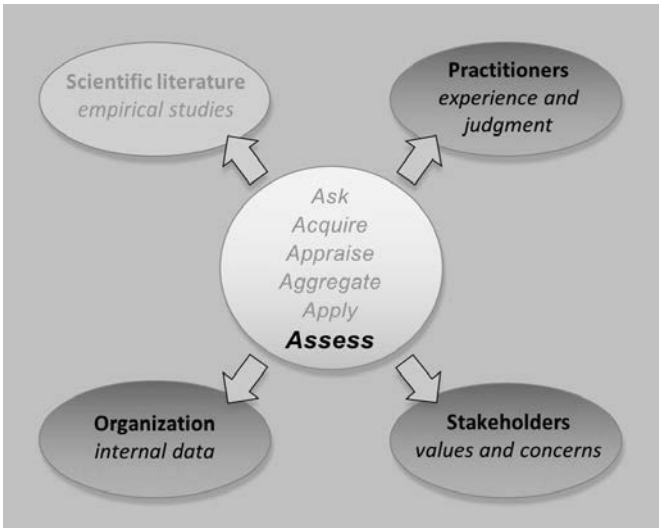
MODULE 14 | ASSESS: Evaluate the Outcome of the Decision

Learning objectives:

- Identify the type of decision (to be) made.
- Determine whether a decision was executed as planned.
- Assess an outcome using the gold standard method.
- Assess an outcome using quasi- or non-experimental methods.
- Determine whether an outcome was measured in a reliable way.
- Assess stakeholder effects.

Introduction

> *You should measure things you care about.*
> *If you're not measuring, you don't care and you don't know.*
>
> *– STEVE HOWARD*



Example

In the United States, the number of employees who work from home has tripled over the past 30 years. Many people still picture an employee working from home as a person in pajamas watching videos on his or her laptop. Several empirical studies, however, suggest that this picture is not accurate and that remote working can have a positive effect on performance. (1)

For this reason, James Liang, CEO of Ctrip, China's largest travel agency with more than 16,000 employees, considered implementing remote working for his call center staff. To make sure his decision had the desired outcome (increased individual task performance), he first assessed the effect of remote working on a smaller scale. A sample of 250 employees were included in the assessment.

Employees with even-numbered birth dates were assigned to the group that worked from home, while those with odd-numbered birth dates remained in the office as a control group. The trial lasted three months. Did the employees working from home resist staying in bed watching TV instead of doing their job? When Liang and his colleagues reviewed the outcome, they were stunned. "It was unbelievable. The work-from-home employees were far from goofing off – they increased their performance by 13.5 percent over those working in the office." In addition, employees who worked from home also reported shorter breaks and fewer sick days and took less time off. Liang decided to introduce remote working for all call center workers. (2)

Assessing the outcome of the decision taken is the sixth and final step of the evidence-based process. After we have ***asked*** critical questions, ***acquired***, ***appraised***, and ***aggregated*** evidence from multiple sources, and finally, ***applied*** the evidence to the decision-making process, we now need to assess the outcome: Did the decision deliver the desired results? Unfortunately, many organizations fail to evaluate the outcome of their decisions – sacrificing one of the principal means to lea n. Remember, the expertise you develop as a practitioner depends in large part on the accuracy of the feedback you get on your decisions – and how well you learn from it, as we discussed in the module *APPRAISE Evidence from Practitioners*.

Systematically assessing the outcome of a decision is one of the key ways in which we can improve the quality of our decisions, and it often leads us to challenge our assumptions, change our judgment, and reconsider our conclusions. For this reason, assessing the outcome of our decisions is something we can and always should do, even – or particularly – in cases in which the preceding five steps of the evidence-based process yield little or no quality evidence. This may be the case, for example, with the implementation of new technologies or in an organizational context that changes rapidly. In those cases, we have no option but to systematically evaluate the results of our decisions through a process of constant experimentation, punctuated by critical assessment of what worked and what did not.

Types of decisions

In this module, we present several methods for assessing the outcome of a decision. Before such an assessment can take place, however, it is important to remind ourselves of the types of decisions managers make, because to a large extent this determines how we can assess the outcome in an effective and reliable way. Following is a brief overview of the common decision types that were presented in the module *APPLY: Incorporate Evidence into the Decision-Making Process:*

- **Routine Decisions**

  Routine decisions, which are often operational, are those that are made within an organization's existing set of rules and procedures. How much of a given product to stock in a shop? Who to hire to fill an existing role? How to deal with a customer complaint? Typically, we can assess the outcome of routine decisions using organizational data generated by systems available within the organization.

- **Nonroutine Decisions**

  Nonroutine decisions involve making changes to the way in which an organization operates. The scale of this type of decision can vary from the introduction of autonomous teams within a particular department to major strategic decisions such as whether to take over another company or launch a new product. The key point about decisions that fall into this category is that they concern something that is new to the organization – but not to the industry – and are thus subject to varying degrees of uncertainty regarding the outcome. Bear in mind that for nonroutine decisions, we may not have standard ways to assess the outcome because existing organizational data are insufficient

- **Novel and/or Hypercomplex Decisions**

  Novel or hypercomplex decisions involve interventions that are not only new to the organization but also new to the industry at large (e.g., introducing an innovative IT solution or starting a business unit in an emerging market). This type of intervention is typically subject to a high degree of uncertainty and risk regarding the outcome and usually involves many unknowns. As a result, we often do not have standard ways to assess the outcome.

Learn by doing 14.1

Read the following scenarios and determine what type of decision/intervention it concerns.

1. *An investment bank that wants to develop and implement a new compensation system for its traders, brokers, and asset managers.*

   a) Routine

   b) Nonroutine

   c) Novel/hypercomplex

2. *An investment bank that wants to build a culture of corporate social responsibility.*
   a) Routine
   b) Nonroutine
   c) Novel/hypercomplex

3. *An investment bank that wants to reduce the absenteeism of its staff.*
   a) Routine
   b) Nonroutine
   c) Novel/hypercomplex

Did I get this 14.1

Read the following scenarios and determine what type of decision/intervention it concerns.

1. *An academic hospital that wants to improve its recruitment and selection procedure of medical staff.*
   a) Routine
   b) Nonroutine
   c) Novel/hypercomplex

2. *An academic hospital that wants to introduce autonomous nursing teams.*
   a) Routine
   b) Nonroutine
   c) Novel/hypercomplex

3. *An academic hospital that wants to use artificial intelligence to reduce surgical infections.*
   a) Routine
   b) Nonroutine
   c) Novel/hypercomplex

Assessing the outcome: Two preliminary questions

In this module, we present several methods for assessing the outcome of a decision. Before we can assess the outcome, however, we must answer two questions: (1) Was the decision executed? and (2) Was the decision executed as planned?

Question 1: Was the decision executed?

Decisions may be made but simply not executed. In an organization that produces physical products or where the decision-makers are in the same building as those executing the decision, it may be possible to check whether a decision has been implemented through "management by walking around." In many organizations, those who execute the decision are located in different departments, buildings, or even countries. In that case, it is worth checking whether the decision indeed has been executed. Even where a decision is clearly represented in new rules, procedures, or training, it does not automatically follow that the decision has been carried out.

Example


News Opinion Sport Culture Lifestyle More

The Guardian

Europe's biggest banks fined for money laundering

All five of the UK's largest banks have been sanctioned for money laundering offences within the last decade

The executive board of an international bank decided that any bank account that has two or more Suspicious Activity reports (reports of activities likely to indicate the account had been used for illegal activity) should be closed. The bank's top executives, however, repeatedly ignored internal warnings that the firm's monitoring systems were inadequate and that it was thus unclear whether the decision was fully implemented. During an audit by the US Justice Department several months later, it indeed turned out that the decision was not implemented by one of the bank's Latin American subsidiaries. This became a major factor in the bank being fined more than $1 billion US in a money-laundering case.

Question 2: Was the decision executed as planned?

This question pertains to implementation fidelity, which describes the degree to which a decision was executed as intended by the decision-makers. It acts as a potential moderator of the effect of the decision and its intended outcomes. Put differently, the degree to which a decision or intervention is executed determines whether the decision produces the desired outcome. Interventions in areas such as medicine and engineering often have dramatic consequences if they are not carried out as specified (people die, bridges collapse). This is one of the reasons we always need to assess implementation fidelity. Unless we make such an assessment, we cannot determine whether a lack of impact is due to poor implementation or the decision itself, and consequently, we may draw false conclusions about the decision's effectiveness.

Example

Tuberculosis is one of the top 10 causes of death worldwide: in 2016, more than 10 million people fell ill with tuberculosis, and 1.7 million died from the disease. For patients with a specific type of tuberculosis, antiretroviral therapy has been shown to have good results. Because the impact of implementation fidelity on the therapy's outcome was unknown, a controlled longitudinal study was conducted. The outcome demonstrated that the overall mortality risk was 12.0 percent. However, under complete implementation fidelity, the mortality risk was 7.8 percent, suggesting that one-third of the mortality is preventable by implementation fidelity. (3)

Implementation fidelity is commonly described in terms of three elements that we need to assess. (4)

1. Content

The content of the intervention concerns its "active ingredients": the actions, knowledge, or skills that the intervention seeks to deliver to its recipients. For example, a company can decide that the leadership training for its managers should include general management skills (e.g., goal setting, performance appraisal, time management) as well as interpersonal skills (e.g., listening, questioning, negotiating).

2. Dose: Duration, frequency, and coverage

The dose of the intervention refers to its duration, frequency, and coverage (i.e., the percentage of the population involved in the intervention). For instance, in the example above, it could be decided that the training should take three days, be repeated every four years, and be mandatory for all middle managers. If an implemented decision adheres completely to the content, frequency, duration, and coverage as determined by the decision-makers, then we can say that fidelity is high

3. Moderators

A high level of implementation fidelity is not easily achieved. In fact, several factors may influence or moderate the degree of fidelity to which a decision or intervention is implemented. For example, the intervention may b very complex and entail multiple elements, the people who are supposed to implement the intervention may lack the necessary skills to do so, or the departments involved may have insufficient financial or human resources to carry out implementation. In those cases, it is likely that implementation fidelity is low.

Did I get this 14.2

Read the following scenario

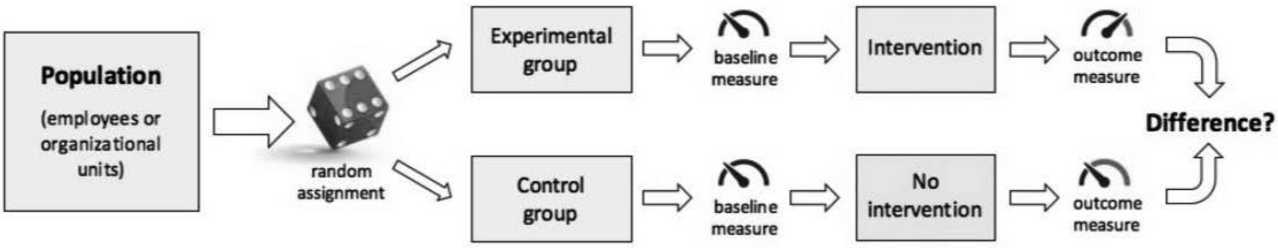
A large manufacturing company decides to implement self-managed teams in which team members are given the latitude to jointly decide how their work is to be done. Until now, the company deployed closely managed teams in which team members have little discretion over tasks and are told not only what to do but how to do it. A pilot that was conducted in three of the company's production facilities found an 18 percent increase in labor productivity after the introduction of self-managed teams. However, six months after the self-managed teams were implemented, it turned out that in 12 of the company's 89 production facilities, the productivity didn't change, whereas in the other 67 facilities, productivity increased between 10 and 20 percent.

What could be the two most likely reasons that the productivity in 12 of the company's 89 production facilities did not change?

a) This could just be coincidence.
b) Self-managed teams may have little or no impact on labor productivity.
c) The managers at the 12 facilities may have lacked the necessary skills to implement the self-managed teams.
d) The 12 facilities may not have implemented self-managed teams.

Assess: The gold standard

When we assess the outcome of a decision, we are asking whether the decision had an effect on a particular outcome. This type of question (Does X have an effect on Y?) is about cause and effect. Questions of this sort are hard to answer. Controlling the independent variable (cause/intervention) and separating it in time from the dependent variable (effect/outcome) can be very difficult. In addition, the e are many confounding factors that may influence the outcome. When it comes to assessing the effect of a decision in a valid and reliable way, we need both a control group (preferably randomized) and a baseline measurement. As you may recall from modules *A Short Introduction to Science*, and *APPRAISE Evidence from the Scientific Literature*, this type of design is known as a randomized controlled trial, considered the gold standard to answer cause-and-effect questions.



A randomized controlled trial (RCT) can be conducted quite easily. For example, in the case of a new service, you can offer the service to a random group of customers, offer no (or a different) service to a control group, and then compare the results. Another example of an RCT is provided at the beginning of this chapter: employees of a call center were randomly assigned to an intervention group (working from home) and a control group (working from the office). The outcomes of RC s are simple to analyze, and the data are easily interpreted. Another domain in which it is relatively easy to run an RCT is marketing, especially online marketing. For example, it is quite common to test the effectivity of online advertisements or web pages by randomizing potential customers to one of the two versions to determine which one performs better. This latter example is also called *A/B testing*.

Not all experiments are created equal!

In science, an RCT is sometimes referred to as a *true experiment*. In daily life, an experiment means trying something out to see what happens and is synonym for *pilot*. A true experiment (RCT) and a pilot, however, are two very different things. A true experiment is a research design with a high internal validity, whereas a pilot means trying out something new (e.g., a new product or service) to determine whether it can be introduced more widely.

Most organizations run pilots (a well-known CEO of a large multinational once complained that his company had more pilots than British Airways), but not many organizations run true experiments. This means that when an organization says it has done an experiment, you should check whether it involved a premeasure, a control group, and random assignment.

In some industries, running experiments (RCTs) is already standard operating procedure. If you order something at Amazon, you are almost certainly part of an experiment – testing products, prices, and even book titles. CapitalOne, an American bank, conducts tens of thousands of experiments each year to *improve the way in which the company acquires customers, maximizes their lifetime value, and terminates unprofitable accounts*. (5)

Baseline

As explained in Module 5, we tend to automatically infer causality because of our System 1 thinking (i.e., automatic, intuitive, heuristical). To determine whether a decision has an effect on a desired outcome, however, we first need to know what the situation was before the decision was executed. Thus, we need a baseline. In athletics, a baseline is essential. Without being sure that all athletes start at the same time (and thus have a similar baseline), we can't be certain that the winner was indeed the fastest. But a baseline is also essential in the realm of management and organizations.

For example, in order to assess the effect of a new working procedure on employees' job satisfaction, we first need to know what their job satisfaction was before the new working procedure was implemented. Here too we need a baseline: a measurement of the metrics of interest before the decision was executed that we can use to compare to later measurements in order to assess the effect. A baseline is also often referred to as a ***before*** measurement, whereas the one taken at the outcome is referred to as an ***after*** measurement.

A baseline is essential not only to determine whether a decision has caused an effect but also to determine what causes what. For example, when the organizational data demonstrates a strong correlation between the level of supervisory support and subordinates' productivity, we are inclined to assume that supportive supervisors cause productive subordinates. However, without a baseline, we cannot rule out what is known as *reverse causality*: subordinates may be productive because they have supportive supervisors, or supervisors may be supportive because they have productive subordinates. (6) When we don't have a baseline, we can't say with certainty whether it was the decision that caused a desired outcome.

Control

In an RCT, we expose a group of employees or organizational units (sometimes called an *intervention* or *experimental* group) to an intervention. This intervention or experimental group is compared with another group that is not exposed to the intervention (sometimes called the *control* or *comparison* group). The control group serves as a benchmark for comparison against the intervention group. In the example above, we could obtain more confidence about the effect of supportive supervisors on subordinates' productivity by comparing the productivity of teams working under a supportive supervisor with those working under an unsupportive supervisor. To rule out alternative explanations for the increased productivity, however, we should assign subordinates or supervisors to the control and intervention groups at random.

Randomization

Even when we use a control group, we can't be 100 percent certain that there were no (unknown) confounders affecting the outcome. The gold standard of outcome assessment therefore includes *random assignment*, a method to create control and intervention groups that include subjects (people [often called *study participants*] or organizational units) with similar characteristics, so that both groups are similar at the start of the intervention. The method involves assigning subjects to one of the groups at random (by chance, like the flip of a coin), so that each subject has an equal chance of being assigned to each group, and any possible distorting factor is equally spread over both groups. Thus, we can more confidently attribute any differences in the desired outcome to the decision.

Example

A large Dutch food retailer is looking for ways to cut its operating costs. The chief operating officer suggests opening all stores one hour later on workdays. The company's executives are divided on the decision's potential impact: some argue that reducing the operating hours will result in a substantial decrease in sales, while others claim that the impact will be minimal. The company's board therefore decides to first assess the impact of the reduced operating hours by conducting an RCT. A sample of 100 stores is randomly assigned to the intervention (reduced operating hours) and the control condition (normal operating hours). After the trial has run for three months, the results demonstrate that reducing operating hours by opening one hour later on workdays does not result in any meaningful sales decline. The board therefore decides to implement later opening hours for all 1,200 stores. (7)

Random sampling is not the same as random assignment!

Random ***sampling*** refers to selecting subjects (people, organizational units, etc.) in such a way that they represent the whole population, whereas random assignment deals with assigning subjects to a control group and intervention group in such a way that they are similar at the start of the intervention. Put differently, ***random sampling*** (or random ***selection***) ensures high ***representativeness***, whereas ***random assignment*** ensures high ***internal validity***.

Learn by doing 14.2

Read the following scenario and determine whether the outcome of the decision/intervention was conducted according to the gold standard.

A large manufacturing company decides to implement self-managed teams. The outcome of the decision was assessed by comparing the labor productivity of the company's 89 production facilities one month before and two months after the implementation. The assessment showed that productivity had increased by 10 percent.

Check all that apply.

a) No, because there was no control group.
b) No, because there was no baseline.
c) Yes, the assessment was conducted according to the gold standard.
d) No, because there was no pre-measure.

Learn by doing 14.3

Read the following scenario and determine how the validity and reliability of the outcome assessment could be improved.

An insurance company decides to implement agile working, a new way of working that focuses less on where or when people are doing work and more on how well they are doing it. Two months after the implementation, the outcome was assessed through comprehensive interviews with nine team leaders and focus groups with 62 employees from eight departments. The outcome indicates both the leaders and the members are very positive about agile working.

Check all that apply.

a) By using a pre- and a post-measure.
b) By using a control or comparison group.
c) No need to improve; assessment was conducted according to the gold standard.
d) By using a post-measure.

Did I get this 14.3

Read the following scenario and determine whether the outcome of the decision/intervention was conducted according to the gold standard.

A large American hotel chain with more than 570 hotels and resorts in 85 countries is looking for ways to improve customer satisfaction. The chief hospitality officer suggests enrolling all frontdesk employees in a two-day training course: Customer Service – Exceeding Expectation. The company's executives are divided on the potential impact of the course. Some argue that it is difficult to train a person to be more service oriented because it's a matter of personality traits such as politeness, empathy, and emotional intelligence: either you have them or you don't. The company's board therefore decides to first assess the impact of the training by conducting a pilot.

A random sample of 50 frontdesk employees from 15 hotels is assigned to the intervention (the two-day training course) and a control condition (two days off). After the pilot, the results demonstrate that the two-day training course had only a minimal effect on the hotels' customer satisfaction scores. The board therefore decides not to implement the training course.

a) No, because there was no baseline.
b) No, because there was no control group.
c) No, because there was no random assignment.
d) Yes, the assessment was conducted according to the gold standard.

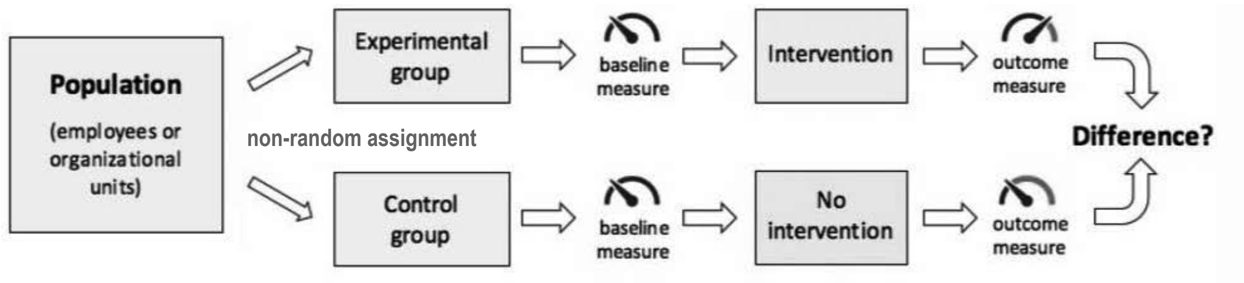
Assess: The silver standard

The gold standard – a randomized controlled trial – requires random assignment of individuals to units (teams, departments, etc.) that serve as either an intervention or a control group. In most organizations, however, units were set up long before the intervention started, and the individuals within these units were often not assigned at random. An alternative option would be to randomize units rather than individuals, but that is feasible only for corporations with multiple, geographically dispersed sites, such as banks, chain stores, government agencies, and health-delivery organizations. (8) In addition, nonroutine change interventions often concern new tasks and responsibilities, but people's tasks and responsibilities are never assigned at random. In these cases, the gold standard is not feasible, so we have no option but to go for silver instead.

Quasi-experiments

The biggest barrier to the gold standard is random assignment. For this reason, we must settle for a nonrandomized controlled outcome assessment, also referred to as a *quasi-experiment*. What makes quasi-experiments experimental is their use of both a control group and a before and after measurement. What makes them "quasi" is the lack of random assignment of people or units to the control and intervention group.

The lack of random assignment negatively affects the assessment's internal validity (trustworthiness). It is nevertheless worth doing them. In fact, in domains such as medicine, many findings a e based on nonrandomized studies, especially when objections of an ethical nature come into play (e.g., in research into the effects of smoking on the development of lung cancer, it would not be ethical to induce one group of people to start smoking in order to see its health consequences). Nonrandomized outcome assessments too can lead to robust empirical insight into a decision's effects, when the assessments are repeated under varying conditions. (8)



Example

A financial services organization with 40 branches in the United Kingdom decides to set goals to enhance the performance of its mortgage sales agents. Because several scientific studies suggest that providing rewards can strengthen a person's goal commitment, which results in better performance, the company also decides to introduce performance-based rewards. To assess the effect of this decision, the company measures the performance of 124 sales agents over a two-month period.

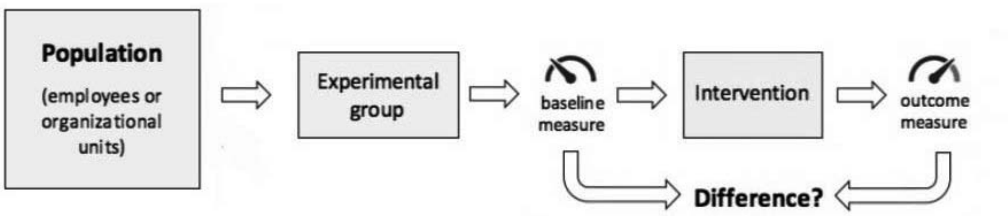
Because it is not possible to randomly assign the agents to one of the two conditions (i.e., since agents often work side by side and would learn, and possibly resent, differences in their reward allocations), 82 agents at six branches serve as the intervention group, and 42 agents at four branches serve as a *control. The key performance indicator is the number of new mortgages sold. The average number of sales in the preceding year is set as a baseline level, and the agents receive a reward when attaining a particular level of performance: 105 percent, 110 percent, and 125 percent, with rewards of $250, $700 and $1,500, respectively.*

At the beginning of the assessment, each agent chooses one of the performance goals and receives the corresponding award if the goal is achieved. Agents at three branches receive cash, while agents at three other branches receive points that are redeemable for rewards listed in a catalogue. After two months, the outcome is assessed: the agents who set goals and received performance-based rewards sold on average 12.5 percent more mortgages. However, it was also found that the agents who were offered cash rewards tended to set higher goals than those who were offered nonfinancial rewards, and as a result, their performance was higher. (9)

Before–after assessment

When random assignment is not feasible, it is important that the subjects (people, organizational units) in the intervention group are similar to those in the control group. If the groups are not similar at baseline, the outcome of the assessment may be flawed. In large organizations with multiple branches, offices, or departments, it is often possible to find a control group that matches the subjects in the intervention group.

In smaller organizations or those with a diverse population, however, a control group may be hard to achieve. The same is true for novel and hypercomplex decisions and for large-scale change interventions. In those cases, it may be preferable to assess the outcome of the decision by only comparing the baseline with the outcome. This type of assessment is referred to as ***before–after measurement***. A before–after measurement may also be more suitable for companies that prefer to implement the decision within the whole organization rather than only in the units that were assigned to the intervention group.



Example

A daycare center has a clearly stated policy that children must be picked up by their parents by 5 p.m. Nevertheless, some parents are often late, and as a result, at least one teacher must wait until the parents arrive. Because there is extensive research suggesting that people are sensitive to financial disincentives, the daycare center manager decides to introduce a fine for tardy parents.

To determine whether this decision has the desired effect, the center keeps track of the number of parents who come late for four weeks both before and after the fine is introduced. In the first four weeks they register, on average, six late pickups per week. In the fifth week, the fine is introduced: parents who arrive more than 10 minutes late will receive a fine of $3. This fine will be added to the parents' monthly bill (which is about $380).

To the manager's great surprise, in the four weeks after the fine is introduced, the average number of late pickups goes up from six to twenty. Apparently, the fine backfired. To find an explanation for this unexpected outcome, a focus group is held with some of the parents. From this focus group, the manager learns that putting such a small fine on a late pickup absolved the parents of the moral guilt they felt for being late. (10)

A before–after measurement is a simple and practical method of assessing the outcome of a decision. Because this type of assessment lacks both random assignment and a control group, however, the outcome is more prone to bias and confounders. For this reason, it is important that we use valid and reliable outcome measures (discussed on the next page) to assess the effect.

Did I get this 14.4

In the example of the daycare center (see above), the manager used only a before–after measurement. Although the outcome is pretty clear, what would be the BEST way to improve the reliability of the assessment?

a) Use a random sample of the parents rather than the whole population.
b) Conduct a focus group before the decision was implemented.
c) Use a control group to see whether the late pickups would also decrease without a fine
d) Randomly assign the parents to an intervention (fine) and control (no fine) group.

Learn by doing 14.4

Read the following scenario

A software development company specializing in mobile apps is faced with a growing number of complaints because of the large number of bugs in the latest release. According to the quality-control managers, the bugs result from coding errors. After a thorough examination of the underlying cause, it becomes clear that errors are being made because most software coders work in an environment where they are easily distracted. Given that evidence from the research literature confirms that even fairly simple distractions tend to increase mistakes eightfold, the company's director decides to take measures to reduce workplace distractions, such as introducing no-interruptions zones. To determine if this decision would indeed reduce the number of coding errors, a pilot is conducted.

What would a quasi-experimental assessment of the pilot's outcome look like?

a) Measure the number of coding errors two weeks before the start of the pilot and two weeks after the end, and determine whether there is a difference.
b) Take a random sample of the company's software coders and randomly assign them to an intervention group (new workplace with no distractions) and a control group (current workplace with the usual distractions), and determine whether the number of coding errors is lower in the intervention group.
c) Take a random sample of the company's software coding teams and assign them to an intervention group (new workplace with no distractions) and a control group (current workplace with the usual distractions), and determine whether the number of coding errors is lower in the intervention group.
d) Take a random sample of the company's software coding teams and assign them to an intervention group (new workplace with no distractions) and a control group (current workplace with the usual distractions), then measure the number of coding errors two weeks before the start of the pilot and two weeks after the end and determine whether there is a difference.

After-action review

When we do not (or cannot) obtain a baseline, it is hard to assess the outcome of a decision. This is often the case in large-scale interventions, hypercomplex decisions, or change projects that have multiple objectives. But even in those cases, we should first make an attempt to assess the outcome, for instance by means of an after-action review (AAR). First used by the US Army on combat missions, an AAR is a structured, reflective evaluation of a recent set of decisions in order to evaluate their effectiveness. Meta-analyses have found that, when appropriately conducted, AARs can lead to a 20 to 25 percent average improvement of the desired outcome (e.g., performance, safety, attitudes). (11)

Nowadays, AARs are used in many disciplines, such as medicine, policing, education, and aviation. The method is relatively straightforward: a facilitator leads individuals or teams through a series of questions that allow participants to reflect on a recent decision and uncover lessons learned in a non-punitive environment. The process may be formal or informal and may last for minutes or hours, but the review always revolves around the same four questions:

1. What did we decide to do?
2. What actually happened?
3. How/why did it happen?
4. What should we do next time?

According to US Army guidelines, roughly 25 percent of the time should be devoted to the first two questions, 25 percent to the third, and 50 percent to the fourth. (12) In addition, it was found that AARs are most effective when the following requirements are met: (13)

- ***Developmental intent:*** The focus of the review should be on learning and improvement rather than evaluation or judgment. A developmental, nonpunitive focus not only yields more honest and accurate feedback but also enhances experiential learning.
- ***Focus on specific events:*** The review should focus on specific activities, episodes, or events rather than on performance or results in general.
- ***Multiple evidence sources:*** The review should be informed by a variety of perspectives and evidence sources. For example, the review should include input from multiple participants and at least one additional source of evidence (e.g., organizational data).

Learn by doing 14.5

Read the following scenario

A global telecommunications company has decided to launch a start-up offering low-cost mobile services to millennials. The new company is offering young customers unlimited data and messages at a cost far below the monthly charge that the average consumer pays for mobile services. Six months after the launch, however, only a few thousand customers have signed up for the new service, mostly people aged 40 and over. To assess the decision to launch the start-up, the CEO decides to conduct an after action review (AAR). All participants in the AAR agree that the start-up's performance is not meeting performance goals. They also agree that the decision to launch the start-up was right but that the poor performance is due to the company's marketing strategy – the marketeers have done insufficient marketing research, and as a result, the wrong customer segment (millennials) was targeted. Based on the outcome of the AAR, the company decides to hire a well-known marketing consulting firm to develop a new marketing strategy.

What are the <u>two</u> main weaknesses of how the AAR was conducted.

a) The CEO failed to discuss the review's purpose and rules.
b) The focus of the review was on performance evaluation and judgment rather than learning and improvement.
c) The review focused on specific activities, episodes, or events rather than on performance and results in general.
d) The review should be informed by multiple evidence sources rather than only the perspectives and judgment of the participants.

Outcome measures

Reliability

When we assess the outcome of a decision, we want to make sure that our conclusions are valid and reliable. For this reason, we prefer to assess the outcome with before and after measurement and a (randomized) control group. The trustworthiness of an assessment, however, is first and foremost determined by the way in which the outcome was measured. The measurement of direct/objective outcomes (e.g., production error rate, staff turnover rate) is more likely to be valid and reliable than that of self-reported/subjective outcomes (e.g., perceived error rate). A detailed overview of aspects to consider when assessing the reliability of an outcome measure is provided in the module *APPRAISE Evidence from the Scientific Literature*, “How Were the Findings Measured?”

In addition, when we assess the outcome of a decision by using organizational data generated by the company’s systems, we need to check whether these data are accurate and reliable, and thus we must consider all aspects that were described in the module *APPRAISE Evidence from the Organization*. For this reason, we strongly recommend completing these two modules.

Did I get this 14.5

A company decides to allow staff to bring their dogs to work, because it is believed this has a positive effect on stress. The outcome of this decision (less occupational stress) can be assessed in different ways. Rank order the four measurement methods based on their reliability (the most reliable method on top).

A. **Stress Observation Scale (SOS)**. Stress on the SOS is defined as the harmful physical and emotional response that occurs when there is a poor match between job demands and the capabilities, resources, or needs of the worker. The scale is filled out by a third person who observes whether a participant is stressed.

B. **Salivary cortisol**. Cortisol is well known to increase with stress and thus is regarded as a highly reliable measure of occupational stress. A single salivary cortisol sample taken at awakening has been shown to be as good as taking multiple samples throughout the day in providing an indicator of overall cortisol production.

C. **Occupational Stress Diary (OSD)**. The OSD is a qualitative measure that provides insight into the way people experience stress. The OSD uses a free response format and is filled out by the employer at the end of the working day. The OSD aims to foster individual reflection and enhance coping skills

D. **Stress Visual Analog Scale (SVAS)**. Stress on the SVAS is defined as one's response to demanding or unpleasant stimuli or conditions. The scale is 15 cm long and anchored at each end with descriptors of “none” to “the most severe imaginable.” The SVAS is a widely used self-report measure with high levels of compliance and acceptable reliability.

Performance measures

When assessing the outcome of a decision, we often have many types of measures available. One of the most widely used measures is performance. Organizations use various methods to measure performance. These methods differ in terms of complexity and are often expressed in a metric that contains both objective and subjective measures. Measuring performance, however, is difficult because it depends on what is defined a *performance* by the organization and its customers' perception of this performance. In addition, the correlation between subjective and objective performance measures tends to be rather low and therefore cannot be used interchangeably. We therefore recommend you use primary outcome measures rather than indirect or aggregated performance metrics. Following is an overview of common primary outcome measures.

Primary outcome measures		
Number of sales	Net profit margin	Job satisfaction
Number of units produced	Return on investment	Staff turnover rate
Number of production stops	Cost/benefit ratio	Absenteeism
Failure frequency rate	Overhead ratio	Retention
Production hours	Market share	Professional time utilization
Throughput time	Company value growth	Research and development quota
Occupancy rate	Customer satisfaction	Timeliness
Waiting times	Number of complaints	Revenue per employee
Number of innovations	Net promoter score	Profit per full-time equivalent employe
Number of patents	Brand awareness	Overtime per employee

Did I get this 14.6

Rank order the four performance measures based on how reliably they can be measured (the most reliable measure on top).

A. The absenteeism rate of nurses in a university hospital.
B. The customer/patient focus of a local hospital.
C. The number of hip replacements performed by an orthopedic surgeon per week in a university hospital.
D. The job satisfaction of nurses in a local hospital.

Costs and benefits

Even the best decision may come with unexpected costs, so we should make a thorough cost/benefit analysis part of every assessment. Several analytics tools and templates are (often freely) available on the Internet. Many of these tools, however, do a poor job of identifying indirect and intangible costs (e.g., a decrease in customer satisfaction or drop in employee morale). Thus, when you conduct a cost/benefit analysis, you need to consult multiple sources of evidence (organizational data, professionals, or stakeholders) to identify all the costs, financial and otherwise

Learn by doing 14.6

Read the following scenario

A French airline company decides to introduce a new policy to reduce costs. The new policy requires customers to pay a surcharge for carry-on items and to pay for meals and seat selection. In addition, one less flight attendant will be assigned in economy class. Six months after the introduction of the new policy, the outcome is assessed. According to the company's management, the new policy is a success because it has produced a cost reduction of 8 percent.

Do you agree that the new policy is a success? Check all three that apply.

a) Yes, because a cost reduction of 8 percent is a great improvement.
b) No, because it is unclear what the impact is on customer satisfaction.
c) No, because it is unclear what the impact is on job satisfaction and employee turnover.
d) No, because it is unclear what the impact is on the number of passengers carried per month/per year.
e) No, because the costs of recruiting and training new staff is not taken into account.

Assessing stakeholder effects

Decisions have both intended and unintended consequences. We are more likely to recognize the latter when we assess evidence from stakeholders regarding a decision's impact. In fact, stakeholders are an important source of information regarding issues we might need to manage in the aftermath of a decision. Unfortunately, some managers prefer to ignore the impact of their decisions on stakeholders, particularly those at lower levels or outside the firm. (14) This evaluation avoidance, however, undermines understanding of the full array of effects of organizational decisions.

This is particularly a problem in nonprofit organizations where positive results of interest to donors and funding agencies are emphasized, while negative effects on clients and the community are downplayed. (15) In for-profit organizations, evidence from stakeholders can inform decision-makers about the effects of their decisions on sustainability, community well-being, and longer-term social consequences.

Note that the values and concerns of stakeholders (in terms of decision outcomes they view as important) can be quite different from what managers believe them to be. For example, in global firms operating in developing countries, locals may be particularly interested in job security and less accepting of cuts that are explained by market factors. (16) Effective (and ethical) decision-making therefore includes systematic assessment of (potential) effects on stakeholders broadly.

Types of stakeholders

Stakeholders take a variety of forms. The most common distinction is probably between *internal stakeholders* (e.g., employees, managers, and board members), *connected stakeholders* (e.g., customers, suppliers, distributors, financiers, and shareholders), and *external stakeholders* (e.g., regulators, government, professional bodies, local communities, and society at large).

Another relevant distinction is that between *direct* and *indirect* stakeholders, reflecting whether a decision has a direct impact on the stakeholder's interests or an indirect impact through the actions of other stakeholder groups. For example, a call center agent working for a retail company that decides to open all stores one hour later on workdays is not directly affected by this decision, but he or she may be confronted by customers who are. This makes the agent an indirect internal stakeholder and the customer a direct external stakeholder.

There is also a distinction between ***primary*** and ***secondary*** stakeholders, which is based on the company's responsibility toward the stakeholder. For example, current employees, customers, and suppliers are primary stakeholders, whereas future employees, regulators, and the local community are secondary stakeholders.

Learn by doing 14.7

Read the following scenario

After a country announced economic austerity measures to control a widening budget gap, it approaches a well-known consulting firm to measure the public reception of these measures. The consulting firm is well aware that the country's government is known worldwide for its authoritarian rule of law and disregard of human rights, but the firm nevertheless decides to accept the assignment.

The consulting firm's analysis shows that the measures received twice as much coverage on social media as in the country's traditional news media and that negative sentiment far outweighed positive reactions. It was also found that only a few social media accounts were driving the negative sentiment. An overview of these accounts was provided in the firm's report.

When the consulting firm evaluated its decision to accept the assignment, it was acknowledged that working for a government that disregards human rights may have a negative impact on the firm's reputation. The counterargument was that the firm is well known for its work with many governments, and it did not engage in any work that sought to target individuals.

Are there any stakeholders (people who may be affected by the firm's practice) you feel are missing in the firm's assessment?

Finally: Keep it simple!

The perfect is the enemy of the good. This old adage applies to many situations, not least of all to assessing the outcome of a decision. As we stated earlier, when the gold standard – a randomized controlled trial – is not feasible, we should go for silver instead: a non-randomized study, a before–after measurement, or an after-action review.

In general, it pays to conduct assessments that are easy to execute and that use minimal resources and staff. Running complex trials that take several months to execute is probably not the best strategy. Instead, we recommend focusing on small-scale assessments that can be automated through the company's existing information systems. Much of what you can learn from large-scale experiments you can also learn from a series of smaller tests.

In addition, sometimes an honest and open after-action review of a failed decision generates more value than a rigorous randomized controlled trial of a successful decision—provided that the organization is open to learning. The goal is not to conduct the perfect assessment – the goal is to learn and help the organization make better decisions.

As Eric Anderson, professor at Kellogg School of Management, rightly states, "What's surprising is not how bad decisions typically are, but how good managers feel about them. They shouldn't – there's usually a lot of room for improvement." (17)

We started this module by reminding you to consider the type of decision that is being made because, to a large extent, this determines how you can assess the outcome in a valid and reliable way. In general, there are three types of decisions:

1. **Routine decisions**: Operational decisions made within an organization's existing set of rules and procedures. Outcomes often can be assessed in a standard way by using evidence generated by systems available within the organization.
2. **Nonroutine decisions**: Decisions regarding changes to the way in which an organization operates. Outcomes often cannot be assessed in a standard way because existing evidence is insufficient
3. **Novel/hypercomplex decisions**: Interventions that are new not only to the organization but also to the industry. Outcomes cannot be assessed in a standard way.

Next, we discussed two questions that must be answered before we can assess the outcome of a decision:

1. ***Was the decision executed?***
2. ***Was the decision executed as planned?***

The second question pertains to *implementation fidelity*, which describes the degree to which a decision was executed as intended by the decision-makers. Implementation fidelity is commonly described in terms of th ee elements that we need to assess:

- **Content**: Its "active ingredients": the actions, knowledge, or skills that the intervention seeks to deliver to its recipients
- **Dose**: Duration, frequency, and coverage
- **Moderators**: Factors that may influence or moderate the deg ee of fidelity to which a decision o intervention is implemented

We then discussed the gold standard for assessing whether a decision or intervention had an effect on a particular outcome: a randomized controlled trial. Controlling the independent variable (cause/intervention) and separating it in time from the dependent variable (effect/outcome) can be very difficult. For this eason, a randomized controlled trial includes three elements:

1. **A baseline measure**
2. **A control or comparison group**
3. **Random assignment**

Quasi-experiment

In many situations, the gold standard is not feasible, so we must go for the silver instead. When random assignment to a control or comparison group is not possible, we must settle for a nonrandomized controlled study, also called a *quasi-experiment*. We explained that nonrandomized outcome assessments too can lead to robust empirical insight into a decision's effects.

Before–after measurement

In smaller organizations or those with a diverse population, a control group may be hard to achieve. The same holds true for novel and hypercomplex decisions and for large-scale change interventions. In those cases, it may be preferable to assess the outcome of the decision by simply comparing the baseline with the outcome. This type of assessment is referred to as a *before–after measurement*.

After-action review

When we do not (or cannot) obtain a baseline, it is hard to assess the outcome of a decision. This is often the case in large-scale interventions, hypercomplex decisions, and change projects that have multiple objectives. But even in those cases, we should attempt to assess the outcome, for instance, by means of an after-action review (AAR). An AAR always revolves around the same four questions:

1. ***What did we decide to do?***
2. ***What actually happened?***
3. ***How/why did it happen?***
4. ***What should we do next time?***

In addition, we explained that AARs are most effective when the review (1) focuses on learning and improvement rather than on evaluation or judgment, (2) focuses on specific activities or events rather than on performance or results, and (3) is informed by a variety of perspectives and evidence sources.

Regardless of the type of outcome assessment, its trustworthiness is first and foremost determined by the way in which the outcome was measured. In general, direct/objective outcomes (e.g., production error rate, staff turnover rate) are more likely than self-reported/subjective outcomes (e.g., perceived error rate) to be valid and reliable. In addition, we explained that measuring performance is difficult because it depends on what is define as *performance* by the organization and on how its customers perceive this performance. For this reason, we recommend using primary outcome measures rather than indirect or aggregated performance metrics.

We also explained that even the best decision may come with unexpected costs, so we should make a thorough cost/benefit analysis part of every assessment. Of the several analytics tools and templates available (often freely) on the Internet, many do a poor job of identifying indirect and intangible costs.

Finally, we discussed that decisions have both intended and unintended consequences. We are more likely to recognize the latter when we assess evidence from stakeholders regarding a decision's impact. As a rule, effective (and ethical) decision-making includes a systematic assessment of (potential) effects on stakeholders broadly.

We ended this module with an important takeaway: Keep it simple! After all, the goal is not to conduct the perfect assessment – the goal is to learn and help the organization make better decisions.

Podcast: ASSESS: Evaluate the Outcome of the Decision

In this podcast host Karen Plum discusses with Eric Barends, Managing Director, Center for Evidence-Based Management , Denise Rousseau,
H J Heinz University Professor, Carnegie Mellon University, Stefanie Nickel, Global Head of Diversity & Inclusion, pharmaceutical manufacturer Sandoz, and Niamh McNamara, Global Head of People & Organisation, pharmaceutical manufacturer Novartis, how to evaluate the outcome of the decision we've taken or the solution we implemented. This is the last stage of our evidence-based management process and is vital to ensuring that we learn from what we've done.

Assessing outcomes is vital, because otherwise, how do we know if what we did was effective, and how can we learn and develop our approach to decision making? Did we capture a baseline before the decision was implemented? And was the decision implemented as planned? If we assess outcomes without these two, then conclusions could be very suspect.

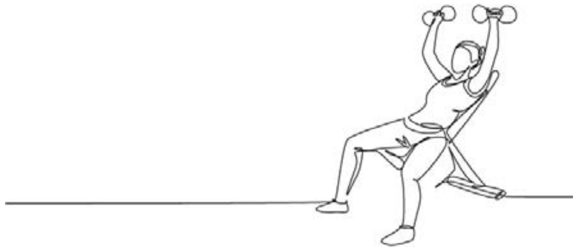
We continue our case study of the large trial at pharmaceutical organization Sandoz (part of Novartis Group), and find out how the D&I interventions we e assessed.

Karen Plum   Niamh McNamara   Denise Rousseau   Eric Barends   Stephanie Nickel

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com

Exercises



Exercise 14.1: Outcome assessment – retrospective

Think about a project, change intervention, or management/business/policy decision you have been involved in (or have observed closely). This project/decision should be reasonably important for your organization, involving significant resources, several or many people, and with no 'easy' answer. Write your answers to the following four questions in detail.

1. What was the outcome the project/intervention/decision aimed to achieve?
2. How was the outcome assessed – what variables or KPI's were measured and how?
3. Is this a reliable way to evaluate the outcome? Explain why / why not?
4. What could have been done to increase the reliability of the outcome evaluation?

Submit a Word document with your answers (max. two pages) according to instructions from your professor/instructor.

Exercise 14.2: Outcome assessment – prospective

Think about a project, change intervention, or management/business/policy decision you are (or have been) involved in. This project/intervention/decision should be reasonably important for your organization, involving significant resources, several or many people, and with no 'easy' answer. Write your answers to the following three questions.

1. What is the outcome the project/intervention/decision aims to achieve?
2. Describe the BEST – that is, the most valid and reliable – way to assess the outcome. Assume you have unlimited resources and the full cooperation of all stakeholders. Describe in detail:
   a) What outcome variables you would measure and how (method, scale).
   b) How you would obtain a baseline measure.
   c) How you would conduct the outcome measure (time, frequency, sample size).
   d) What your criteria for success (or failure) would be.
3. In most organizations only limited time and resources are available and the best way to assess the outcome may not be possible. Therefore, describe the MOST FEASIBLE – that is, given the available resources and cooperation, the most practical, realistic, yet informative- way to assess the outcome. Describe in detail:
   a) What outcome variables you would measure and how (method, scale).
   b) If (and how) you could obtain a baseline measure.
   c) How you would conduct the outcome measure (time, frequency, sample size).

Submit a Word document with your answers (max. two pages) according to instructions from your professor/instructor.

Suggestions for further reading

- *After-Event Reviews: Drawing Lessons From Successful and Failed Experience*, Ellis & Davidi, JAP 2005
- *Treat Your Organization as a Prototype: The Essence of Evidence-Based Management*., Jeffrey Pfeffer & Robert I Sutton. Design Management Review, 2006
- *A conceptual framework for implementation fidelity*., Carroll et al, Implementation Science, 2007
- *A step-by-step guide to smart business experiments*., Anderson et al, HBR, 2011
- *Do Team and Individual Debriefs Enhance Performance? A Meta-Analysis*, Tannenbaum & Cerasoli, Human Factors, 2013
- *Field experiments in organizations*, Annual Review of OPOB, 2017
- *Guide To The After Action Review*, Salem-Schatz, Center for Implementation Practice and Research Support, 2010

CEBMa center for Evidence-Based Management

Open Learning Initiative

References

1. De Menezes, L. M., & Kelliher, C. (2011). Flexible working and performance: A systematic review of the evidence for a business case. *International Journal of Management Reviews*. Volume 13. Number 4.
2. Example adapted from Surdoval, A.(September 20, 2017). *Why working from home should be standard practice*. https://ideas.ted.com/why-working-from-home-should-be-standard-practice/.
3. Patel, M. R., Westreich, D., Yotebieng, M., Nana, M., Eron, J. J., Behets, F., & Van Rie, A. (2015). The impact of implementation fidelity on mortality under a CD4-stratified timing strategy for ant etroviral therapy in patients with tuberculosis. *American Journal of Epidemiology*. Volume 181. Number 9. 714-722.
4. Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*. Volume 2. Number 1. 1-9.
5. Anderson, E. T. (2011). A step-by-step guide to smart business experiments. *Development and Learning in Organizations: An International Journal*. Volume 25. Number 6.
6. Bass, B. M., & Bass, R. (2008). *The Bass handbook of leadership: Theory, research, and managerial applications*. The Free Press. New York.
7. Example adapted from Thomke, S., & Manzi, J. (2014). The discipline of business experimentation. *Harvard Business Review*. Volume 92. Number 12. 70-79.
8. Eden, D. (2017). Field experiments in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*. Volume 91-122. 91-122.
9. Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons. Hoboken, NJ.
10. Presslee, A., Vance, T. W., & Webb, R. A. (2013). The effects of reward type on employee goal setting, goal commitment, and performance. *Accounting Review*. Volume 88. Number 5. 1805-1831.
11. Levitt, S. D., & Dubner, S. J. (2014). Freakonomics: *A rogue economist explores the hidden side of everything*. Morrow. New York.
12. Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors*. Volume 55. Number 1. 231-245.
13. Garvin, D. A. (2003). *Learning in action: A guide to putting the learning organization to work*. Harvard Business Review Press. Boston, MA.
14. Keltner, D., Druenfeld, D. H., & Anderson, C. (2003). *Power, approach, and inhibition*. Psychological Review. Volume 110. Number 2. 265-284.
15. Julian, R. (2016). Is it for donors or locals? The relationship between stakeholder interests and demonstrating results in international development. *International Journal of Managing Projects in Business*. Volume 9. Number 3. 505-527.
16. Lähteenmäki, S., & Laiho, M. (2011). Global HRM and the dilemma of competing stakeholder interests. *Social Responsibility Journal*. Volume 7. Number 2. 166-180.
17. Anderson, E. T. (2011). A step-by-step guide to smart business experiments. *Development and Learning in Organizations: An International Journal*. Volume 25. Number 6.

Learn by doing & Did I get this? Answers and Feedback

Learn by doing 14.1

1a,c: Incorrect

1b: Correct

Developing and implementing a new compensation system is most likely “new” to the bank – but not to the industry. This means that we may not have standard ways to assess the outcome

2a,b: Incorrect

2c: Correct

The evidence on this topic is very limited and probably low in quality, so this type of intervention involves many unknowns. It should therefore be regarded as a hypercomplex decision for which we have no standard ways to assess the outcome.

3a: Correct

3b,c: Incorrect

Reducing absenteeism is an issue every organization deals with on a regular basis and for which much evidence is available. Consequently, we most likely can assess the outcome using organizational data already available within the organization.

Did I get this? 14.1

1a: Correct

1b,c: Incorrect

The recruitment and selection of personnel is an issue every organization deals with on a regular basis and for which much evidence is available. Consequently, the hospital most likely can assess the outcome of the improvement using organizational data already available within the organization.

2a,c: Incorrect

2b: Correct

Introducing autonomous teams involves making changes to the way in which an organization operates. It is something that is new to the organization, though not necessarily to the industry or other firms. This means the e may not be standard ways to assess the outcome.

3a,b: Incorrect

3c: Correct

Introducing an innovative solution that is relatively new to the industry is subject to a high degree of uncertainty and risk regarding the outcome and involves many unknowns. As a result, there is often no standard ways to assess the outcome.

Did I get this? 14.2

a,b: Incorrect

c,d: Correct

It is unlikely that self-managed teams may have little or no impact on labor productivity, as the pilot found a rather large positive effect. In addition, it is unlikely that coincidence or chance can explain why the productivity in the 12 facilities did not change: the difference between those 12 and the other 77 facilities is just too large.

Learn by doing 14.2

a: Correct

b,c,d: Incorrect

Learn by doing 14.3

a,b: Correct

c,d: Incorrect

Did I get this? 14.3

a,b,d: Incorrect

c: Correct

Did I get this? 14.4

a: Incorrect > Using a random sample would decrease the reliability of the assessment. It is always better to include the whole population.

b: Not quite right > After all, it is unclear whether the parents would admit in advance that they were happy to pay a small fine for a late pickup

c: Incorrect > You are correct that using a control group would be a good way to improve the assessments reliability, but there is a better answer.

d: Correct

Learn by doing 14.4

a: Incorrect > This assessment lacks a control group and is therefore a before–after assessment, not a quasi-experimental assessment.

b: Incorrect > This assessment includes random assignment, which makes it an experimental assessment, not a quasi-experimental assessment. An experimental assessment is the gold standard, but random assignment often is not feasible, so we have no option but to go for a quasi-experimental assessment.

c: Incorrect > This assessment lacks a premeasure, so it is not a quasi-experimental assessment.

d: Correct > This is what a quasi-experimental outcome assessment looks like.

Learn by doing 14.5

a,c: Incorrect

b,d: Correct

Did I get this? 14.5

1: B

2: D

3: A

4: C

In general, the measurement of direct/objective variables (in this case, taking a sample from someone's saliva, which is then sent to a laboratory) is more likely to be valid and reliable than a self-reported measurement, so we would rank this method on top.

In addition, symptoms of stress are hard for a third person to observe. Also, it is stated that the SVAS is "a widely used self-report measure with high levels of compliance and acceptable reliability," but no information on the SOS's reliability is provided, so we would rank the SVAS higher than the SOS

Finally, although no information on the reliability of the SOS and OSD is provided, the OSD is a qualitative method used to foster reflection, so we would rate this as less eliable method than the SOS

Did I get this? 14.6

1: C

2: A

3: D

4: B

Both outcome A and B can be considered direct/objective variables, but given the number of observations involved, the number of hip replacements performed by one orthopedic surgeon can be measured more reliably than the absenteeism rate of a large number of nurses. Job satisfaction and in particular customer focus are harder to measure, so we would rank these measures as the least reliable.

Learn by doing 14.6

a,e: Incorrect

b,c,d: Correct

The new policy may not be welcomed by the company's passengers and thus may have a negative impact on customer satisfaction. Ultimately, this could result in a drop in the number of tickets sold/ number of passengers carried. In addition, the flight attendants may feel that educing the number of crew members increases their workload, which may have a negative impact on job satisfaction and employee turnover.

Learn by doing 14.7

The scenario above is actually a real life example. The consulting firm in this scenario was McKinsey & Company, and the country involved was Saudi Arabia. After McKinsey issued its report, several critics were arrested and imprisoned by the Saudi government, and a large number of accounts were shut down. Apparently, the consulting firm did not realize its report could be misused by the Saudi government and failed to assess (in advance) the possible negative effects on an important group of stakeholders: people who are critical about the government's austerity measures. As a result, McKinsey's reputation was seriously damaged.


The New York Times

Wednesday, October 24, 2018

McKinsey Takes Heat for Its Saudi Work

McKinsey & Company was in the spotlight on Monday after The New York Times reported that the consulting firm had prepared a report that may have played a role in Saudi Arabia's effort to target its online critics.

MODULE 15

Building the Capacity for Evidence-based Management

Introduction

> *Be the change that you wish to see in the world.*
>
> *– MAHATMA GANDHI*

NOTE: This module does not contain '**Learn by doing**' or '**Did I get this**' exercises.

In 1954, Leon Festinger, an American social psychologist, read about a cult that believed the end of the world was nigh. The cult was led by 'Mrs Keech', a housewife from Chicago who claimed to have received a message from extra-terrestrials that on December 21 a great flood would extinguish all life on earth, although all 'true believers' would be rescued the night before by a flying saucer. Mrs Keech attracted a large group of followers who gave up their jobs, sold their homes and gave away all their possessions. Given the followers' strong beliefs, Festinger wondered how they would react when Mrs Keech's claim would turn out to be false. To find out, he posed as a believer and infiltrated the cult. At midnight on December 20 the evidence was clear: no spaceship had turned up. The followers, however, concluded that this time the world had been spared because of the faith of their group. The next day they actively started to recruit new members. (1) (2)

We like to think that we base our judgments on evidence. After all, isn't this the hallmark of rationality? Of course, as explained in the module *APPRAISE – Evidence from Practitioners*, our judgment is prone to cognitive biases that negatively affect the decisions we make. But when confronted with hard, undisputable evidence that reveals we should change our judgment, we do just that. Or do we? Unfortunately, as the example above demonstrates, simply showing the evidence is often not enough to change a person's beliefs. (3) We could argue that the cult members did not change their minds because they were dumb or foolish, and that this example doesn't represent the average manager. That argument, unfortunately, doesn't ring true. People's beliefs and judgments – including those of managers and business leaders – are part of a complex system that is resilient or stubborn (take your pick) in the face of contradictory evidence. As a result, simply showing your boss, colleagues or employees strong evidence may not be enough to change what they believe. In fact, as illustrated by the following experience of Jeffrey Pfeffer, renown management professor at Stanford University, managers at times even choose to ignore the best available evidence. (4)

"A few years ago, while serving on the compensation committee for a publicly traded company, we were considering what to do about the CEO's stock options and our stock option program in general. Just that day articles in the mainstream business press were published on research showing that stock options led to risky behaviour. (5) *That research added to the growing body of evidence demonstrating that many executive pay practices not only did not enhance company performance, but led instead to misreporting of financial results.* (6) (7) *At the meeting, a vice president from Aon Consulting who was advising the compensation committee replied "No" without any hesitation or embarrassment when I asked him first, if he knew about this research and second, if he was interested in me sending him the original articles or other information about the extensive research on stock options and their effects. What is particularly telling is that many people from other compensation consulting firms to whom I have related this story said it could have been their firm, too – that the perspective reflected is typical."*

To give evidence-based management a shot at success, we need to increase the capacity of managers and organizations to prioritize quality evidence over personal opinion – and incorporate what the body of evidence indicates into their professional judgment. That, however, is easier said than done. In this module, we will discuss how you can build the capacity for evidence-based management – not only in yourself, but also among your peers, bosses and the organization at large.

Becoming an evidence-based manager

This module's opening quote from Gandhi has energized generations of people to act as change agents for a better world, by first starting with themselves. Evidence-based management also starts with ourselves. By becoming an evidence-based practitioner, you can inspire other practitioners and form the basis of an organizational culture that is itself evidence-based. An approach of this kind is not only about mastering the six skills – Ask, Acquire, Appraise, Aggregate, Apply and Assess – but also the lifelong pursuit of personal and professional development as an evidence-based practitioner. This development typically involves three phases:

1. Developing a questioning mindset
2. Making decisions more explicit
3. Practicing and learning everyday

Developing a questioning mindset

The absence of evidence-based decision-making is not so much due to a lack of knowledge about the evidence supporting or contradicting a decision (though that is a serious issue), but rather the absence of managers with a questioning mindset. (8) Wondering what works, what doesn't and why is the first step towards improving management practice. As discussed in the first modules, asking questions kicks off a deliberate search for evidence and leads to the active exploration of alternatives. This questioning also increases understanding by testing assumptions about problems, solutions and the outcomes of decisions. Developing a questioning mindset that appreciates the difference between trustworthy and less trustworthy evidence is a first step to becoming an evidence-based manager. Developing this mindset, however, is a career- (and maybe even life-) changing proposition.

An evidence-based practitioner thinks differently from other people. Evidence-based management is not just about applying skills and knowledge: it's about taking a different perspective on the organization and its day-to-day concerns. This shift starts with developing the habit of frequently asking yourself and others, "What's the evidence for that?" Of course, raising questions can be scary, especially for those of us who fear making waves. But, once practiced at it, we become comfortable asking, "Is this your personal opinion based on your own professional experience? Is there any other evidence to support it?" This habit of asking questions has turned many evidence-based practitioners into the "evidence squad", as they learn to apply their questioning mind in a manner that promotes raising critical questions without necessarily criticizing. A must here is to learn ways of raising critical questions in socially effective ways (read: civil and persuasive). To be effective, we need to avoid being dismissed as a mere naysayer. In addition, we need to learn when to ask these questions. For example, we should question the evidence regarding an important strategic decision at the beginning of the decision-making process (preferably when the problem to be solved is first defined). I s less effective to wait until it's at the CEO's desk for final app oval. At the same time, don't be afraid to speak up when the available evidence contradicts established practice or political interests. A questioning mindset is the lifeblood of evidence-based management.

Making more mindful and explicit decisions

As a manager or leader, you make decisions all the time. But as an evidence-based practitioner, you should avoid making decisions on auto-pilot. Instead, we need to foster mindful, explicit decision-making. The process of making decisions of this kind has two parts. The first concerns decision awareness, recognizing the numerous daily choices we and our organizations make. Try making a list of the events of a morning or afternoon at work. What situations did you encounter, and what did you do or say? We bet you make far more decisions or choices in a day then you realize. For this reason, evidence-based managers tend to make many decisions with the evidence they already have in hand. Thus, it is valuable to 'prime the pump'. This means acquiring and learning in advance – the evidence useful for decisions you most likely will make. Priming the pump is especially useful for repeat decisions like hiring, running meetings, making purchasing decisions and the like. This can mean reading up on subjects related to important everyday decisions. But what about those decisions you cannot foresee? Of course, evidence-based management is not about taking an evidence-based approach to all decisions. Instead, it means paying attention to the number and type of decisions you make, and being able to recognize when evidence is important to pursue.

The second part of making more mindful and explicit decisions is paying attention to how decisions are actually made. Analyze an important management decision you have been involved in. Ask what was the problem to be solved? From whom or where was the available evidence in this decision obtained? What kind of evidence was available and from which sources? What evidence was used and what not? Did some types of evidence influence the decision more than others? Was evidence missing? What are the indicators that the decision was a success (or a failure)? How long did the whole decision-making process take? Making decisions more mindful and explicit prompts critical thinking and evidence-seeking behavior. It means paying attention to how our decisions might be made differently. Doing so provides you an opportunity to apply the six steps and skills described in this module – and to make decisions using the best available evidence.

Everyday practice and learning

Basing your management decisions on the best available evidence can be a turning point in your career as a manager or business leader. It is a big step, and not always easy. But the more you practice, the better and easier it will be.

As you become more experienced, you will develop more advanced evidence-based skills, such as searching in online research databases, critical appraisal of scientific evidence, or using Bayes Rule to aggregate evidence from multiple sources. Effective practice and learning require lots of repetition and exposure to a variety of different conditions (organizational settings, types of decisions, etc.). But, as explained in the module *APPRAISE – Evidence from Practitioners*, repetition and exposure alone do not necessarily result in greater expertise – a person still needs direct, objective feedback. This means that systematically assessing the outcome of decisions is a key way to improve your skills and knowledge as an evidence-based manager. In the module *ASSESS: Evaluate the Outcome of the Decision*, we described several methods of assessing a decision. As you may recall, the best way to assess a decision's outcome tends to be a systematic before/after comparison and use of a control group, but sometimes a frank and thoughtful After-Action Review generates considerable value – provided we are open to learning from mistakes.

Effective practice, however, not only requires lots of repetition and objective feedback, but also the opportunity to practice using an evidence-based approach. The opportunity to practice means having a work environment that allows use of evidence-based skills. As a rule, it is determined by the amount of discretion and control you exert over how you do your work (i.e., professional autonomy). In a highly structured work setting (i.e., with lots of rules about how a task is done), your opportunity to practice may be somewhat limited. Nonetheless, you still may be able to negotiate with your supervisor for the flexibility to apply an evidence-based approach – perhaps on a project or a change program. A work setting that supports introducing evidence into conversations – at meetings, with staff, managers and clients – is a strong indicator of organizational readiness for evidence-based management.

Building evidence-based capacity among bosses and peers

On the previous page we described activities you can perform by yourself. However, unless you are running a one-person business on a desert island, evidence-based management involves other people, such as colleagues, bosses or clients. Therefore, the next step in building evidence-based capacity is introducing the approach to the people in your organization. In general, bosses and peers appreciate the professionalism and conscientiousness that evidence-based practitioners display. Yet, there still may be pushback, for example when decisions need to be quick or if there's lots of politicking and backbiting in your organization. In such cases, executives may choose to ignore the evidence and make decisions based on personal judgment (*"I don't think we need more evidence – we all know what we need to do, so let's just get it done"*). Even in work environments where evidence use is highly valued, if evidence challenges a boss or peer's closely held beliefs, it can activate defences and muddle their judgment.

Evidence alone doesn't change people's minds

Evidence alone does not change minds – political scientists have demonstrated this in many empirical studies. For example, in 2010 a landmark study showed that confronting people with hard evidence can backfire, making them more entrenched in their biases and misperceptions. (9) The reason for this counterproductive effect is simple: When someone presents us with evidence contrary to our beliefs, we can feel threatened and dig in our heels. Particularly when the evidence makes people feel stupid (i.e., threatens self-perceptions regarding their intelligence and competence), they may put their porcupine needles up and dispute (or ignore) the facts.

When this person is your boss or someone higher up in the organization, the evidence may even be perceived as undermining their formal authority. In addition, the human brain seems intrinsically reluctant to go back on a decision taken, even when the facts underlying the decision have changed. Put differently, we seem to be wired to stick to our judgments and deal with the consequences later. (10) As a result, when presented with contradictory evidence, individuals tend to dig in their heels and increase their commitment to an initial belief. This shouldn't surprise us. After all, many psychological mechanisms in the human mind induce self-justification, confirmation bias, and a host of other self-protective mindsets. So, what can we do instead?

Giving the mind a way out

As a rule, when people feel threatened, their minds dig in rather than give in. The key is to trick the mind by giving it a way out. (11) For example, explain to a person that his/her judgment is right, given what he/she knows. But now new information has emerged – showing that the underlying facts have changed. Thus, it now makes sense that his/her judgment changes too. Keep in mind that the moment you (implicitly) belittle, ridicule or embarrass the other person with contradictory evidence, you've lost the battle. Instead, as Robert Cialdini, Professor of Marketing and Psychology at Arizona State University and recognized "persuasion" expert, explains, you must offer the other person a way in order to save face while getting out of his/her prior commitment: "Well, of course that was your judgment a week ago, because this new evidence was not yet available." (12)

Helping people to develop a new operating logic

A meta-analysis based on a large number of randomized controlled studies found that a well-argued and detailed debunking message was better at persuading people to change their mind than simply labelling their judgment as plain wrong. (13) Stripping out incorrect or unreliable information leaves a painful gap in a person's operating logic regarding how the world works. As a result, contradictory evidence is more effective when it provides information that enables people to develop a new logic or narrative that in turn legitimizes their change in judgment.

Enhancing people's understanding of science

An additional finding of the meta-analysis mentioned above was that enhancing people's understanding of science can increase acceptance of contradictory evidence. Rebutting bad science may not be effective, but asserting the methodological quality of trustworthy scientific evidence seems to be effective. (14) In the module *APPRAISE – Evidence from the Scientific Literature*, we have provided several examples that can be used to educate your bosses and peers on what constitutes trustworthy evidence based on science. Including practical examples of reliable research findings enhances people s understanding even more, especially when doing so invokes an emotional reaction, ranging from fear (wow, cultural diversity can lead to task conflicts and poor communication) to relief (ok, increased absenteeism is not uncommon after a period of restructuring). Lastly, you should word evidence from scientific research in a way that passes the mother-in-law test. That is, you should be able to explain the findings to your mother-in-law. (15)

Increasing people's accountability

A trend promoting evidence-based management is public demand for accountability. Evidence-based medicine is not only about making better clinical decisions – it is also about the need to justify clinical decisions to others. (16) Accountability is a serious issue in management too. Managers, leaders and administrators often endure a great deal of criticism from various directions. Mismanagement, incompetence and misuse of power are the charges most commonly heard. As a result of this increasing social pressure, there is a strong drive for increased accountability. Accountability refers to the (implicit or explicit) expectation that one may be called upon to justify one's beliefs, actions or decisions. It implies that people who do not provide a satisfactory justification for their decisions will su fer negative consequences – ranging from disdainful looks and public outcry to discharge and prosecution. (17) Not surprisingly, a large number of studies indicate that increasing a decision makers felt sense of accountability leads to more information-seeking behaviour, less implicit bias, and greater openness to external evidence.

In addition, managers and leaders who experience increased accountability regarding how they make decisions (process accountability) will be more open to evidence that challenges their beliefs than those who are held accountable only for the outcome of their decisions (outcome accountability). (18) (19) Increasing accountability in order to enhance more systematic evidence use is helped when the parties involved have influence over the decision process, for example as a supervisor, CEO, board member, or stakeholders in a position to publically review decisions (e.g., shareholders or review committees).

Take small steps and pick your battles

It is tempting to try and apply all at once the skills and knowledge this module presents, but instead we advise caution and mindfulness. Evidence-based management challenges existing beliefs and conventional management practice. Your bosses, peers and clients may need time to get used to this new approach. Instead of flooding them with organizational data, stakeholder input and scientific evidence, take small steps and focus on one or two aspects of evidence-based management at a time. You might make a habit of providing bosses and peers with succinct plain language summaries of relevant studies on important issues. You might collect and share relevant organizational data as it becomes available. In doing so, pick your battles. An evidence-based manager need not take an evidence-based approach to all decisions since some battles are not worth fighting – some arguments a e lost before they even begin. For example, some people are so certain of their beliefs that no amount of evidence can change their minds. Those persons are, by definition, fundamentalists, to which an evidence-based argument is doomed to fail. On the other hand, over time an accumulation of wins through big and small uses of evidence to improve decision quality can really pay off.

Building evidence-based capacity in your organization

The final step of building evidence-based capacity concerns the organization at large. The cultural meaning and value of evidence varies across firms, with technical and clinical organizations potentially being more receptive. For example, Google Inc. structures its employee-selection procedures around both research findings and organizational data in order to avoid unconscious bias in its promotion and hiring decisions. (20) However, making evidence-based management organizational and not just personal, involves raising collective awareness about the added value of evidence for management-related decisions. Broadcasting the idea of evidence-based management to the larger organization can involve a variety of interventions – from conversations and lunchtime meetings in which new research findings a e discussed to citations of research in internal memos, reports or policy papers. At the same time, an organization's capacity for evidence-based management is largely determined by its leadership and climate.

Organizational climate and leadership

An essential prerequisite of organization-wide evidence-based management is senior leadership that promotes an EBMgt 'climate'. A global survey suggests that managers tend to have positive attitudes towards evidence-based management, and a large majority believe use of scientific evidence improves the quality of managerial work — while lack of time is commonly seen as a major barrier. (21) You can overcome such barriers by having a senior management that promotes an organizational climate where it is psychologically safe for members to raise concerns about evidence quality and where supports are in place to acquire the best available evidence before making important decisions. This prerequisite is underscored by a recent systematic review finding that supportive leadership and organizational climate are key factors in the implementation of evidence-based management. (22)

Building an evidence-based climate can take a lot of forms. It can be bottom-up, as the distinctive style in which individual managers practice, make decisions and lead their teams. Or it can be top-down, led by the board or senior leadership in order to create a basis of common understandings and shared values ("this is the way in which our company makes decisions"). Finally, it can also be intervention-driven, introducing evidence-based initiatives in a drive for culture change. For example, process changes such as the sustained implementation of after-action reviews can alter several cultural features including values, norms and patterns of behavior. When you promote an evidence-based climate, opportunities for intervention include:

- **Focus of attention**: What does the organization pay attention to, what is measured, and what is controlled? Does the organization focus on sustaining decision quality or on short-term outcomes? Do managers pay attention to developing employees' capacity to think critically and encourage them to use multiple sources of evidence when making decisions?
- **Reactions to crises and critical incidents**: What messages does the organization send when problems arise? In a crisis, employees look to their leaders for signals reflecting the organization s key values. When all eyes are on you, the evidence-based leader has a valuable opportunity to convey the organization's priorities. For example, if a failure occurs, is learning appreciated more than avoiding being caught making mistakes? Do employees see a commitment to evidence-based processes for improving the organization's practices, or do they suspect a cover up or a blame game?
- **Reward systems**: Who is selected? Who gets promoted? Who leaves? How important rewards are allocated can signal the value leaders place on members who follow an evidence-based approach. On the other hand, efforts to promote an evidence-based management climate are undermined if good evidence-based practitioners leave while less conscientious managers are promoted.

- **Modelling**: How might evidence-based management actually look in this organization? Members need to see how people in jobs like their own might take an evidence-based approach. It helps to make visible how managers diagnose problems in an evidence-based fashion, search for evidence, assess outcomes, etc. Importantly, calling attention to the processes through which people make decisions and making them transparent can both educate organization members regarding evidence-based decision making and enhance their trust in the organization.

Organizational resources

Finally, organizations need resources in order to successfully build the capacity to practice evidence-based management. You can think of organizational resources in terms of three categories: Ability, Motivation and Opportunity to Practice.

- **Ability**: Building collective evidence-based skills and knowledge among organization members requires training people in the six skills we have detailed in this book. This training is particularly effective if it starts with senior leadership and then cascades throughout the organization. By training senior leaders first, we take a step toward changing the work environment in the direction of a more evidence-based organizational climate. (23)
- **Motivation**: Creating a critical mass of people who support evidence-based management helps to create new organizational norms regarding evidence use in decision making. Training is a start for developing such norms. The interventions we have described above for culture change are themselves motivation-altering mechanisms. Managers who act as a role-model and reward their employees for using evidence in making decisions are creating new incentives and supports that motivate the organization's uptake of evidence-based management. Importantly, norms that support evidence-based management are enhanced by making it psychologically safe for members to ask critical questions regarding the organization's decisions and practices.
- **Opportunity to Practice**: Opportunity to practice refers to the support that the organization provides to engage in evidence-based management. Examples of such support include staff who can easily search for relevant studies in research databases; information systems that can capture and process organizational data enabled by an IT department that performs basic statistical analyses; HR-systems routinely gathering professional and stakeholder evidence, and HR professionals able to design in-house questionnaires, conduct surveys and run focus groups; staff who can develop reliable outcome measures and conduct RCTs, quasi- experiments or after-action reviews. Some organizations have installed so-called Evidence Assessment Teams, whose members are tasked with obtaining evidence on a practice question or pending decision. Other organizations have appointed a Chief Evidence Office, responsible for ensuring that the company uses the best available evidence to inform its decision makers. All these types of support make a huge difference in the amount of time involved in making evidence-based decisions, and thus substantially increase employees' opportunity to practice evidence-based management.

Some final words

With this module on building evidence-based capacity we have reached the end of this course. We hope the insights, tools, and checklists provided will help you in taking an evidence-based approach to the problems, opportunities and solutions you encounter in your daily practice and help your organizations make better decisions.

This module started with how to become an evidence-based practitioner and ended with how to build the organization's evidence-based capacity. Still, we believe, this course should end with building evidence-based capacity within the society at large. To this end, we would like to draw attention to the words of one of our proof readers: *"What I take from this course is the career- and life-changing value of taking an evidence-based approach, which is fostering an inquisitive mind that appreciates the difference between trustworthy and less trustworthy evidence. It is not only committed to helping managers and organizations making better decisions, but also about making this world a better place."*

Decisions made by managers have a profound impact on the lives and wellbeing of people the world over. As Henry Mintzberg, famous management thinker, once said, *"No job is more vital to our society than that of a manager. It is the manager who determines whether our social institutions serve us well or whether they squander our talents and resources."* By ignoring evidence, billions of dollars are spent on ineffective management practices, to the detriment of employees and their families, communities, and the society at large. As evidence-based practitioners, we have a moral obligation to change this situation. We can do this by helping organizations to find the best available evidence, to critically appraise its trustworthiness, and to encourage critical thinking and dialogue about assumed problems and preferred solutions. Let's not forget that evidence-based practice started as a movement that had ambitions that surpass the realm of individual practitioners and organizations.

As we close, we are reminded of what Amanda Burls and Gordon Guyatt, two of the movement's pioneers, have said, *"… evidence-based practice is not just about changing a person's skills and knowledge, it's giving them a totally different perspective on the world... It's an activist thing. We want them to get out … and change the world."*

Let's get out and do it!

Podcast: Building the Capacity for Evidence-based Management

In this podcast host Karen Plum discusses with Eric Barends, Managing Director, Center for Evidence-Based Management, Denise Rousseau,
H J Heinz University Professor, Carnegie Mellon University, Jonny Gifford, Senior Researcher, CIPD, David Creelman, CEO of Creelman Research, Stefanie Nickel, Global Head of Diversity & Inclusion, pharmaceutical manufacturer Sandoz, and and Steven ten Have, Partner TEN HAVE Change Management & Professor of Strategy and Change at VU University Amsterdam, how to build evidence-based management capacity in organisations.

There are 3 levels – you as an evidence-based manager; your colleagues who you can influence and guide in evidence based approaches; and the skills, practices and processes in your organisation.

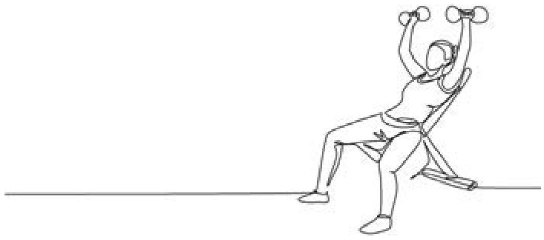
We hear from several experts about ways to build on what you've learned, and how others have approached the development of evidence-based practices in their organisations, including the CIPD, the professional HR body, about how it promotes this practice among its membership.

And you're never alone, there are always others forging a path in evidence-based management who you can reach out to, learn from and share with.

Karen Plum | Jonny Gifford | Denise Rousseau | Eric Barends | Steven ten Have | David Creelman | Stephanie Nickel

Listen on buzzsprout

https://evidencebasedmanagement.buzzsprout.com



Exercises

Exercise 15.1: Guided reflection – becoming an evidence-based manager

Reflection is a powerful process: it creates learning from your experience. Without reflection your experiences may quickly be forgotten and their learning potential lost. By reflecting on the insights and skills you developed in this course and critically evaluating their application in daily practice, you can deepen your understanding of an evidence-based approach. Consequently, you can improve the quality of your decisions. For this reason, we ask you to write a reflection paper in which you reflect on what you have learned in the course. A reflection paper is written in the first person, free flowing, subjective, analytical, and requires a certain time investment. To support your writing, however, you can use the questions below as a guide.

1. What do you consider the *three most important insights* you gained through this course? Explain why. Were they new to you? Did they challenge your existing beliefs or assumptions? Did they affect your daily practice as a manager, leader, consultant, or policy maker? If so, how?
2. What do you consider the *three most important skills* you learned? Explain why.
3. What did you learn from taking an evidence-based approach *in your daily practice* as a manager, leader, consultant, or policy maker? Which of the six steps were easy to apply and which did you find challenging Why?
4. What did you *discover about yourself* as a practitioner/decision-maker? Did the course and the exercises change your professional perspective in any way?
5. What do you consider your greatest strengths as an evidence-based practitioner? What are your *greatest challenges*? Explain why.
6. What are your *goals for the next 6 months* > what are you going to try to do, or do differently?

Submit a Word document with your answers (max. two pages) according to instructions from your professor/ instructor.

Exercise 15.2: Building evidence-based capacity among bosses and peers

As explained in the module, unless you are running a one-person business on a desert island, evidence-based management involves other people, such as colleagues, bosses or clients. Therefore, an important step in building evidence-based capacity is introducing the approach to the people in your organization. For example, you could:

- Invite your colleagues, managers, and other people within your organization who might be interested for an 'awareness session' – an interactive, entertaining but slightly challenging (max. one hour) meeting. Here you can give a presentation about evidence-based management: what it is, why we need it, and why you feel this approach could benefit your organization. Base your presentation on the topics addressed in module 1, or download a PowerPoint slide deck here >>

- Conduct a CAT *(see exercise 6.1. and 7.1)* on a topic relevant to your organization. Present your findings t interested colleagues, managers, and other people within your organization. End your presentation with two or three practical recommendations fairly easy to implement.
- Collect and analyze organizational data on an organizationally relevant topic *(see also exercise 8.1 and 9.1)*. Present your findings to interested colleagues, managers, and other people within your organization. End your presentation with two or three practical recommendations fairly easy to implement.

Choose one of these three interventions. After you have given your presentation, write down your answers to the following six questions:

1. ***How did the presentation go? How did your audience react?***
2. ***What surprised you?***
3. ***What did you find challenging?***
4. ***What questions and/or comments from the audience did you find hard to answer?***
5. ***Do you feel your presentation contributed to a better understanding of the value of an evidence-based approach? Why / why not?***
6. ***What would you do differently next time?***

Submit a Word document with your answers (max. two pages) according to instructions from your professor/instructor.

Exercise 15.3: Building evidence-based capacity in your organization

The final step of building evidence-based capacity concerns the organization at large. Although an organization's capacity for evidence-based management is largely determined by its leadership and culture, there are small, practical and impactful steps you can take. Indeed, this is how evidence-based practice cultures start in organizations.

1. Critically evaluate the evidence-based capacity of your organization. In particular, assess your organization's capacity to *acquire and appraise* evidence from the four sources:
   - *Practitioners* – the professional experience and judgment of practitioners
   - *Scientific literature* – findings f om empirical studies published in academic journals
   - *Organization* – data, facts and figu es gathered from the organization
   - *Stakeholders* – the values and concerns of people who may be affected by a decision
2. Discuss your findings with your boss and/or other stakeholders
3. Write a proposal (for your boss, stakeholders, or other audience) in which you describe in detail the steps that could be taken to improve your organization's use of *one source of evidence*.
4. Alternatively, you might focus your proposal on a *specific consequential or recurrent decision*, and write a proposal in which you describe the steps that can be taken to improve your organisation's use of relevant evidence sources.
5. Alternatively, you might write a proposal to introduce a *specific evidence-based practice* in your organization. Examples of possible practices include After Action Reviews, decision checklists for routine practices, conducting critically appraised topics (CATs) for critical decisions, or decision outcome assessments

Submit a Word document with your proposal (max. two pages) according to instructions from your professor/instructor.

Suggestions for further reading

- *Politics Over Analytics: BI and Data Science Findings Need a Pitch*, Vaughan Robison, LinkedIn blog, January 2015
- *Laszlo Bock – Work Rules Highlights*, London Business Forum, 2015
- *Why we cover our ears to the facts*, Matthew Syed, BBC Magazine, 2016.
- *An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability*, Hall et al, JOB, 2017
- *Why Facts Don't Change Our Minds - New discoveries about the human mind show the limitations of reason*., Elizabeth Kolbert, The New Yorker, 2017
- *Robert Cialdini on how persuasion works in business and politics*, Cardiff Garcia, transcript of interview with Robert Cialdini, Financial Times Alphaville, 2017

References

1. Suls, J. (2017). *Leon Festinger*. Encyclopedia Britannica.
2. Festinger, L. (1956). *When Prophecy Fails*. University of Minnesota Press.
3. Syed, M. (2017). *Chilcot: Why we cover our ears to the facts*. BBC News Magazine.
4. Pfeffer, J. (2012). *The Oxford handbook of evidence-based management*. Oxford university press.
5. Sanders, W. M. & Hambrick, D. C. (2007). Swinging for the fences: The effects of CEO stock options on company risk-taking and performance. *Academy of Management Journal.*
6. Dalton, D. R., Certo, S. T. & Roengpitya, R. (2003). Meta-analyses of financial performance: Fusion o confusion?. *Academy of Management Journal*.
7. Burns, N & Kedia, S. (2006). The impact of performance-based compensation on misreporting. *Journal of Financial Economics*.
8. Kovner, A. R. & Rundall, T. G. (2006). *Evidence-based management reconsidered.* Frontiers of health services management.
9. Nyhan, B. & Reifle , J (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*.
10. Burnett, Dean (2017). *The neuroscience of no regrets: why people still support Brexit and Trump.* The Guardian, Science.
11. Varol, O (2017). *Facts don't change people's minds*. Here's what does. Heleo.
12. Garcia, C (2017). Robert Cialdini on persuasion. *Financial Times.*
13. Chan, M. P. S., Jones, C. R., Hall Jamieson, K. & Albarracín, D. (2017). Debuncking: a meta-analysis of the psycological efficacy of messages countering misinformation. *Psychological Science*.
14. Gawande, A. (2016). The mistrust of science. *The New Yorker*.
15. Eden, D. (2017). Field experiments in organizations. *Annual Review of Organization Psychology and Organization Behavior*.
16. Daly, J. (2005). Evidence-based medicine and the search for a science of clinical care. University of California Press.
17. Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*.
18. Scholten, L., Van Knippenberg, D., Nijstad, B. A. & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology.*
19. See De Langhe, B., van Osselaer, S. M., & Wierenga, B. (2011). The effects of process and outcome accountability on judgement process and performance. *Organizational Behavior and Human Decision Processes.*
20. Bock, L. (2015). *Work rules!: Insights from inside Google that will transform how you live and lead*. Hachette.
21. Denison, D. R. (1996). What is the difference between organizational culture and organization climate? A native's point of view on a decade of paradigm wars. *Academy of management review*.
22. Flodgren, G., Parmelli, E. & Doumit, G. (2011). *Local opinion leaders: effects on professional practice and health care outcomes*. Cochrane Database of Systematic Reviews.
23. Rousseau, D. M. & Gunia, B. (2016). Evidence-based Practive: The Psychology of EBP. *Annual Review of Psychology*.