
1 Introduction: what do we need evaluation for?

Reinhard Stockmann

PROLOGUE

The port of departure was Southampton, the destination is New York. Meanwhile, the liner is a few hundred miles south-east of Newfoundland. Briskly she ploughs her way through the North Atlantic. The night is cold and moonless; the sea lies there with an unaccustomed smoothness, like some fishpond in which all the stars and the gleaming lights from the cabins and dining halls are reflected. Dinner is just being served; the mood is warm and vivacious.

The name of the steamship is *Titanic*. She is regarded as unsinkable and is the world's largest and most luxurious liner. She is on her maiden voyage. There are more than 2200 passengers and crew on board: on the lower decks, emigrant families with their modest possessions; up above, many of the world's wealthiest men and women. There is no hint yet of the drama which will be set in motion by a series of overwhelming nautical errors in the hours to come.

For two days now, the radio operator has been receiving regular reports from other vessels on the positions of icebergs. The meticulous records of the Fourth Officer on the chart show that, slowly but surely, the course of the vessel and the ice are moving closer together.

Sixteen hours before the accident, the ice is only a little way north of the scheduled route; ten hours before, it is practically right on it. At lunch, the captain showed the relevant radio message to the chairman of the board of the *Titanic*'s owners. The latter cast an eye over the telegram and stuffed it into his jacket pocket without so much as a word. He either failed to grasp the meaning of it or it interfered with his intention, officially undeclared, of having the *Titanic* cross the Atlantic in record time on her very first voyage.

At about 9 p.m., the captain, a seaman who began his career on sailing vessels and has been given the maiden voyage of the *Titanic* as his last, honourable assignment, appears on the bridge. The officer on watch draws his attention to the temperature, which has dropped by 6 degrees Celsius to freezing point in the last two hours.

2 *A practitioner handbook on evaluation*

At the same time there is lively Morse traffic in the radio room. The radio operator is intending to take advantage of the proximity of the land station at Newfoundland's Cape Race to send passengers' greetings to their relatives back at home and in turn to receive the latest stock market prices. Being busy with lucrative telegrams from passengers leaves him little time to concentrate on nautical matters. Later investigations will show that further ice warnings received in those final hours never actually reached the bridge at all.

Without any significant change of course and at undiminished speed, the three blades of the enormous propeller drive the ship onwards. The officer on watch does impress on the lookout, stationed constantly in the crow's nest high up above deck, the importance of remaining watchful on account of icebergs. But icebergs, often covered in dark glacial drift, are difficult to make out at night, especially in a calm sea when there are no waves whose phosphorescent ripples forming around the berg would make the latter easier to see. Not only that, aboard the luxury liner *Titanic* the officers on watch on the bridge may well be equipped with binoculars, but the lookouts, who are in a much better position to see, have to manage without.

When the lookout begins to make out a black mass silhouetted against the stars above the horizon, he yanks the bell-rope without a moment's hesitation. But the evasive action now taken can no longer suffice. Like some enormous fingernail, the berg slits open the hull of the ship along a third of its length. It is shortly before midnight on 14 April 1912. The ship, named after those giants of Greek mythology who rebelled unsuccessfully against the gods, slowly begins to tilt over bow first (Figure 1.1), to sink three hours later in the calm sea.

The *Titanic* – though it must be said that this did conform to the regulations which applied at the time – had lifeboats for only half the people on board. Having said that, hundreds of places in them remained empty. While two-thirds of the first-class passengers were saved, only a quarter of those from third class survived. Investigations which followed the accident showed that it was by no means the case that the passengers on the lower decks – as rumours had had it – were prevented from getting into the lifeboats, but that no one had actually told them that the vessel was sinking. So it was that most of those who were emigrating, in spite of the vessel's growing list, remained where they were almost right up until the end, too timid to climb up to safety through the quarters of the upper classes and reluctant to abandon the few possessions they had brought with them for the start of their new life. (Adapted from DEZA, 1997: 'Monitoring – staying in touch with reality'.)



Source: http://commons.wikimedia.org/wiki/File:St%C3%B6wer_Titanic.jpg.

Figure 1.1 'The Sinking of the Titanic', artwork by Willi Stöwer, 1912

ON SHIP VOYAGES, PROJECTS, PROGRAMMES AND EVALUATION

What does this example have to do with evaluation? Well, the ship voyage to New York described here can be understood as a '*project*' which was planned and at the implementation of which an attempt – even if the outcome was catastrophic – was made. The effects caused by this '*project*' can still be felt today.

Indeed, projects and programmes have quite a few things in common with ship voyages. They begin at a certain point in time and pursue a certain aim. The '*route*' by which this aim is to be achieved is determined beforehand.

The *achievement of this aim* is subject to a number of dangers and risks, for example, the rigours of the weather, but the different *interests* of the various individual actors may also prevent it from being achieved. On board ship, for example, there may be a mutiny if insurmountable

differences arise between the crew on one side and the captain on the other. Programmes too can fail because of *differences* between the various actors. Thus it often happens during development that the partners pursue interests and goals which are quite different to those of the people who provide the funds. While some, for example, may be interested primarily in technical equipment and thoroughly modern expertise, others may perhaps merely intend to assist with inexpensive but good advice.

This not only means that voyages, like programmes, require *careful planning*, in which the route and the means by which the destination is to be reached are determined, but also that it is advisable to make sure who is actually pursuing what aims, since aims are not always obvious and may deviate from those which are 'officially' postulated. In the example of the *Titanic*, there may, apart from the officially declared aim of reaching New York, also have been an unofficial aim, namely, that of reaching New York in record time so as to win the coveted 'Blue Riband' on the vessel's very first crossing, with the result that insufficient heed was paid to important information (for example, the ice warnings) and safety precautions neglected.

Qualified personnel are required for the undertaking both of ship voyages and programmes. In principle, this was the case on board the *Titanic*. However, there may be doubts in the case of the captain, of all people, the most important man on the vessel, for he was a quartermaster 'who began his career on sailing-vessels' and was 'awarded the maiden voyage of the *Titanic* as his last, honourable assignment' (adapted from DEZA 1997: 2). It is possible that he was not sufficiently well qualified to assume command of a 22-knot liner.

In order for personnel to be able to work together in a constructive way, an *organizational structure* and clear *task specifications* are required. This too was the case on board the *Titanic*. However, there were some shortcomings in the execution of those tasks. While the lookouts in the crow's nest carried out their duties as conscientiously as they could (not having been issued with binoculars), the radio operator, instead of taking heed of the incoming ice warnings, chose to occupy himself with the stock market prices.

There were problems with *communication* too. The ice warnings were not passed on to the bridge, or were ignored by the captain (telegram, drop in temperature). And above all, the passengers on the lower decks – with the dire consequences now known to us all – were not even told that the ship was sinking! This is cited as the main reason why as many as two-thirds of the first-class passengers were rescued, but only a quarter of those in third class.

For the undertaking of voyages, as for that of programmes, *technologies*

are used. The *Titanic* was the largest and most modern ship of its time, and was even thought to be unsinkable. The use of the most modern technologies, however, can easily lead to an overestimation of what is technically feasible and to a neglect of the risks and dangers. Not only that, but the technology that is available, of course, has to be used properly in order to develop its effects. If, for example, the lookouts in the crow's nest of the *Titanic* had been equipped with binoculars (whereas in fact it was the officers on the bridge who had them), it might have been possible to detect the iceberg earlier.

And last but not least, in order to be able to carry voyages – like programmes – into effect, *financial resources* are necessary too: for the shipping company to build the new ship, for the passengers to buy tickets for the passage. Furthermore, costs play a central role in any voyage and any programme, most of the actors having to husband their resources efficiently and keep a close eye on their expenses and income for that reason. It is possible that the northerly course of the *Titanic*, on which the risk of encountering icebergs is very high, was also followed for reasons of cost, being considerably shorter.

Voyages – like programmes – are undertaken in specific *environments*, which are subject to constant change. Sea voyages may be subject to bad weather and are sometimes threatened by pirates or, as we have seen, by icebergs. This means that however good the planning is, the predetermined course cannot be followed blindly. The example of the *Titanic* makes it all too obvious that proceeding in such a way can lead to disaster. Every ship voyage and every programme requires competent *guidance* in which an eye is kept not only on the course, that is, the route that leads to the destination, but also on the destination itself. It must, for example, be ascertained whether or not a ship (or programme) is still on course, in other words whether or not the destination can still be reached under the given conditions. Apart from that, however, from time to time the destination itself should be called into question. Bad weather, for example, can make it necessary to call at a port other than the one which was scheduled. Economic crisis, war or natural disaster can jeopardize the original aims of a programme, so that it may be necessary to alter them or, conceivably, even to discontinue the programme. Since the environment (the surroundings) is subject to constant changes which cannot be foreseen in the planning phase, the individual factors which influence the reaching of the destination must be kept under constant observation so that management decisions can be made.

This is where the topic of '*evaluation*' comes into play, or rather an instrument closely related to it, '*monitoring*'. While monitoring is designed to ascertain whether or not the scheduled goals of the programme are going to be achieved within the agreed period and whether or not the

destination is being approached in the way that was planned, evaluation also deals with questions which go beyond that, such as whether or not the destination is able to be reached under the given circumstances, or perhaps needs to be modified or even changed completely.

So just as on a ship voyage the sextant enables the captain to ascertain whether or not he is still on course, the instrument of monitoring helps the programme manager to ascertain whether or not the implementation of the various individual steps planned is actually going to lead to the achievement of the aims. With the aid of the instrument of evaluation he can also check (or have someone else check) whether or not the aims which were laid down in the planning phase ought still to be adhered to unchanged.

Results-based monitoring (such as accounting) provides a constant supply of information by comparing what actually is, in other words that which has been performed or assessed, with what was supposed to be, the target values laid down in the planning phase. If the actual situation is repeatedly recorded at intervals, the 'description of the route' can be compared with the 'course' laid down in the planning phase. If the difference between the actual and the planned course is no longer tolerable, there must be a 'course correction'.

Management consists in the observation of processes and results, their assessment, the decisions derived from that assessment and the action taken upon them. However, it is equally obvious that not all factors are manageable. This applies both to internal circumstances which relate to the provider and to external circumstances in the environment of programmes and implementing organizations. The provider, for example, cannot employ better qualified personnel – even if this has been recognized as necessary on account of a central deficiency – if such personnel are simply not to be had on the job market, or if the salary structure of the provider is so unattractive that no one who is qualified actually applies (which is, for example, very often the case with schools in developing countries). Neither are economic crises, disasters or other external events able to be influenced by a project or programme. In those cases, that is just where the art of management lies: in adapting the aims of a project or programme accordingly.

Evaluation not only supplies the kind of data that monitoring supplies relating to the question of whether or not those being evaluated are still on the 'right' course ('Are we doing the things right?'). It also looks into the question of whether or not it is actually the right course ('Are we doing the right things?').

Evaluation, furthermore, is not merely a descriptive activity which simply gathers data in accordance with a predetermined pattern; it also involves an assessment component. This assessment can be made – as will be shown later – by applying some very diverse criteria (the reaching of

the destination being only one among many). Evaluation not only supplies data and assessments relating to the process course of a programme, but also, and above all, data and assessments relating to the *impacts* produced by the programme interventions. These in turn are divided into intended and unintended impacts. Both these forms of impact – as shown in detail in Chapter 2 – are recorded, taken stock of and assessed in evaluations.

The *Titanic* failed to achieve her declared (intended) aim – that of calling at New York. Worse still, her loss had catastrophic (unintended) impacts. Entire families were wiped out, those rescued were traumatized for the rest of their life, capital was destroyed, ocean travel avoided, and so on. On the other hand, the accident may have brought about some positive (unintended) impacts by causing the safety concepts for ship voyages to be revised, and shaking people's blind trust in technology. Even in quite different areas of activity (for example, that of evaluation), lessons were learned from the loss of the *Titanic* . . . and are still being learned today.

The recording and assessment of impacts is one of the most important tasks in programme evaluations. The aspects investigated include the achievement of aims, the benefit reaped by the programme participants, the impacts of the programme on non-participants, and so on. Not only that, but evaluation also has the job of finding out how a given result came about and what *causes* are responsible for the effects observed. In doing this, it is not only the individual programme components and intervention elements that need to be elaborated.

The following also need to be evaluated:

- which (official and covert) aims and so on are being striven towards in the programme
- which interests the different groups of people involved are pursuing
- what degrees of competence and qualification levels those entrusted with carrying out the programme have
- what the functionality of the structures of the implementing organization is like
- how the communication structure and cooperation between those involved in the programme function
- whether or not the technologies being used are suitable and appropriate for the implementation of the programme
- whether or not sufficient financial resources are available for the implementation
- how the properties (structures) and situational conditions of a programme are likely to change in the course of time
- and, finally, whether or not the planning and management procedures employed cater to the complexity and the problems of the

course of the programme; in other words, to what extent the programme management is functional.

ANSWERS TO THE QUESTION ‘WHAT FOR?’

Taking the example of the *Titanic*, we can come up with several answers to the question we asked at the beginning, ‘*What do we need evaluation for?*’

Some useful insights for shipping were able to be gained from the evaluation (that is, the analysis and assessment) of the loss of the *Titanic*. Investigations undertaken *ex post* like this enable integral, so-called summative contemplations, in which the intended and unintended impacts which came about can be taken stock of and assessed in their entirety.

Furthermore the decisions, activities, processes initiated and outputs achieved which led to certain results and effects can be analysed. The causality question – which results and effects were brought about by which causes? – also plays a decisive role.

Not only that, the example of the *Titanic* has made it clear that evaluation and the instrument with which it is ‘twinned’, monitoring, are of decisive importance for current undertakings and programmes, because they provide and assess information which puts management decisions on a rational basis. Conducted *ex ante*, that is, before the voyage or programme has begun, evaluation can also be used in the planning phase, checking the feasibility of the aims and ways in which those aims can be achieved, and assessing them.

Thus evaluation is used for purposes of *rational management*, but also for *continuous learning*. When, for example, deficiencies and problems in implementation are revealed and risks and dangers pointed out, process sequences can be improved and errors avoided. In this more formative perspective of current programmes, the learning processes can be used directly for the (re-)shaping of those programmes. Particularly in the case of evaluations which are designed *ex post* and are thus automatically summative, it is no longer the direct design that predominates, but learning for the future. From the *Titanic* catastrophe, for example, some useful lessons were able to be learned for the shipbuilding industry, safety technology and equipment, route planning and management, and so on, though travellers did not come to benefit from this until later. An important prerequisite was that the results were not kept secret but made public. Thus not only were the *Titanic*’s owners able to benefit (that is, learn) from this, but so were the shipping sector and, even, the general public.

Here we come across another answer to the question ‘*What for?*’ Evaluation can also serve a purpose of ‘*enlightenment*’. By showing how

successfully, effectively, sustainably, and so on programmes have gone, who has benefited from them and who has not, what mistakes were made and should be avoided in the future, and so on, evaluation creates *transparency*. In the example of the *Titanic* it was possible to determine which mistakes led to the catastrophe and who was responsible for them. In the evaluation of political programmes, for example, it becomes clear whether or not they have made a significant contribution to the solution of existing social problems, how their costs and benefits are proportioned, whether or not the policy-makers have kept their promises, and so on. When governments, authorities or non-governmental organizations (NGOs) avail themselves of evaluations to document (or have someone else's document) the performances they have achieved (output), the aims they have accomplished (outcome), the effects that have been produced altogether (impact) and how long-lasting those effects are going to be (sustainability), the *legitimacy* and *credibility* of political measures can also be underpinned.

This means that there is not just one answer to the question 'What do we need evaluation for?' Evaluation can serve a multitude of purposes, which is systematized still further in Chapter 2. But we can already make a note of the fact that evaluation, founded on the systematic gathering, analysis and assessment of data:

- can make a contribution to rational management
- taps sources of learning in order to develop measures and programmes further
- creates transparency by providing information on implementation processes and the achievement of aims and effects
- and thus also makes a contribution to the assessment of the legitimacy of measures and programmes.

In its time, the inquiry into the loss of the *Titanic* not only contributed to people's understanding of how the catastrophe came about so that mistakes made could be recognized, lessons for shipping learned, the general public informed about the accident and the responsibilities revealed; even today, the *Titanic* catastrophe can still serve to make it clear to what ends evaluation can be used. The remaining chapters of this book are intended to show in detail how this happens.

AIMS AND STRUCTURE OF THIS BOOK

In order for evaluations to achieve their greatest possible usefulness, people should be aware of what makes a '*good*' evaluation and how it

should be planned and conducted. This knowledge must of course be possessed by those who are going to conduct the evaluation, and also by those who commission it. Clients should at least know what purposes they are using the evaluation for, what use they can put the evaluation findings to and what quality features distinguish a professionally conducted evaluation.

This is where this book starts from. *Its aim* is to provide an overview of how the various instruments of evaluation can be deployed for their main tasks in the planning and implementation of measures and programmes and the assessment of their results and effects. Basic knowledge and practical recommendations for the application of evaluation procedures and methods are imparted in a concise and user-friendly way. The book draws upon the contents and documentation from numerous introductory and training courses on evaluation at the Center for Evaluation (CEval) and those used in the creation of monitoring and evaluation systems. The book is aimed at those who as yet have but little knowledge of evaluation and would like to continue their education in this field.

Chapter 2 offers some basic specialized knowledge of evaluation. This includes putting evaluation on the map of the realm of empirical social science, the systematization of its various aims, a synoptical presentation of the many different kinds of evaluation approach and the presentation of an example of an impact-orientated evaluation model, ending with a section which deals with some of the fundamental issues relating to the conducting of evaluations.

Chapter 3 discusses evaluation approaches which compete with each other and which complement each other. One by one, instruments such as controlling, the balanced scorecard, benchmarking and auditing, which have been enjoying growing popularity for some time now in the non-profit sector too, are compared with evaluation. By elaborating the similarities and differences between these concepts, which were developed in business management, and the evaluation approaches which have their origin in social science, the book makes it clear that the various different concepts can also complement one another.

While these two chapters pursue the aim of providing a conceptional overview, the chapters that follow impart not only a good basic knowledge but also practical recommendations for the application of evaluation procedures and methods, and draw attention to the special contextual features of evaluation as compared with those of scientific fundamental research.

Chapter 4 describes how an evaluation is conducted from an organizational point of view by going through the individual planning and implementation steps from the drawing up of a tender to practical management

tasks, using a practical example. All the subsequent chapters go more deeply into topics which those conducting evaluations need in detail in the course of their work.

First in the preparation of an evaluation, for example, is how to gather the data on which the object being evaluated (a measure, project or programme) is to be assessed. Various evaluation designs for this are presented in Chapter 5. Even if these are, to a great extent, identical to conventional research designs, evaluation confronts design development with some special challenges, since the clients, and in some cases also the evaluatees, are involved in the process.

Chapter 6 concentrates explicitly on the social context of evaluation, which differs considerably from the traditional research context. Evaluators can adopt different roles in evaluations, from which implications arise for the process itself, for the relationships between the evaluators and others involved (stakeholders), for the qualification requirements the evaluators are expected to meet and for the usefulness of evaluations.

As we mentioned at the beginning, the criteria with which an evaluand is assessed are of decisive importance for the results of that assessment. Different criteria examine a circumstance from different perspectives. In order to be able to assess whether or not certain criteria have been fulfilled, 'measurements' need to be made. What measuring means in a social science context and how, with the aid of indicators, circumstances which are not able to be recorded directly can be rendered measurable are the themes dealt with in Chapter 7.

In order to be able to carry out 'measurements', in other words in order to ascertain which values a given indicator has produced, 'data collection' is necessary. Which methods present themselves for data collection and which special data collection circumstances need to be taken into account in evaluation are the main subjects of Chapter 8.

If various data collection methods are used in order to record as many aspects of an evaluand as possible, and in order to improve the validity of the evaluation findings, a large quantity of (quantitative and qualitative) data are 'produced', and it is necessary to organize these data to suit the questions. Data which are not analysed in this way do not produce any benefit and end up in 'data cemeteries'. Effective data management ensures that the data are analysed purposefully and efficiently. How this works and what needs to be taken into account when doing it are the subjects of Chapter 9.

The final chapter of this book (Chapter 10) covers the topic which also makes up the final section of each and every evaluation: 'reporting'. This comprises not only the final report, in which the most important findings, conclusions and recommendations of an evaluation are presented,

but also the presentation of those findings to clients or other groups of people involved (for example, the evaluatees). Having said that, reporting relates not only to the concluding phase of an evaluation, but also to all the reporting and presentation activities during the course of an evaluation, from the very first meeting for the clarification of the assignment to the final editing of the final report or the writing of a press release to the general public.