Evaluating Environmental Education

Stokking, H., van Aert, L., Meijberg, W., Kaskens, A.

IUCN Commission on Education and Communication CEC

IUCN
The World Conservation Union

Evaluating Environmental Education

IUCN – The World Conservation Union

Founded in 1948, IUCN – The World Conservation Union brings together States, government agencies and a diverse range of non-governmental organizations in a unique world partnership: over 900 members in all, spread across some 139 countries.

As a Union, IUCN seeks to influence, encourage and assist societies throughout the world to conserve the integrity and diversity of nature and to ensure that any use of natural resources is equitable and ecologically sustainable.

The World Conservation Union builds on the strengths of its members, networks and partners to enhance their capacity and to support global alliances to safeguard natural resources at local, regional and global levels.

IUCN Commission on Education and Communication – CEC

The Commission on Education and Communication is one of IUCN's six Commissions, a global network of voluntary, active and professional experts in environmental communication and education, who work in NGO, government and international organisations, professional networks and academic institutions. Commission members' daily work is about how to encourage people to take responsibility in their personal and social behaviour for the environment. CEC specialists are experts in learning processes, stakeholder management, how behaviour is changed and in communication management.

CEC advocates the importance of intergrating communication and education in environmental projects and programmes. CEC promotes more strategic use of communication as a policy tool and builds capacity to better plan and use education and communication for IUCN's mission.

For CEC communication and education are the means to have biodiversity, environmental and equity concerns a part of people's decision making.

Evaluating Environmental Education

Stokking, K., van Aert, L., Meijberg,W., Kaskens, A.

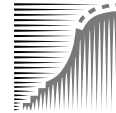Translated from Dutch by Rosemary Martin

The designation of geographical entities in this book, and the presentation of the material, do not imply the expression of any opinion whatsoever on the part of IUCN or other participating organizations concerning the legal status of any country, territory, or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The views expressed in this publication do not necessarily reflect those of IUCN or other participating organizations.

The text of this book is printed on offset A7, tolally chlorine-free paper

Contents

Foreword

Environmental education is a must and so is the evaluation of environmental education.

There are at least four good reasons to devote due attention to evaluation:
• it clarifies what pupils and adult participants learn from nature and environmental education activities;
• it creates opportunities to account to funders about performance and effectiveness;
• it provides information improving the quality of future educational activities and consumer satisfaction;
• it promotes professionalism in nature and environmental education.

For this reason the Netherlands Ministry of Agriculture, Nature Management and Fisheries financed the development of suitable evaluation methods for nature and environmental education. These methods were originally prepared in a three volume handbook in Dutch.

Now in partnership with IUCN - The World Conservation Union's Commission on Education and Communication we bring you this edited version of *Evaluating Environmental Education* in a more accessible language, English.

By developing this international version we want to make the material available to a wide group of users, from those working in field study centres, protected areas, World Heritage Sites or educational institutions around the world.

For IUCN and the Ministry of Agriculture, Nature Management and Fisheries – a member of IUCN – the conservation of nature in all its biological diversity is a priority. Both organisations believe that conservation and equitable and sustainable use of nature are a basis for sustainable development. The mission statement of the Nature Management Directorate is "nature for a sustainable society". The vision of IUCN is "a just world that values and conserves nature." Environmental education is an important tool in achieving these goals.

We express the hope that this handbook will be widely used. Interested users can get an overview of the work, learn systematically about various options for evaluation, or target the detailed information they require. There is no doubt that this would contribute to achieving the high quality we seek in the ways people take responsibility for nature and in the professional delivery of environmental education.

Johan F. de Leeuw
Director General of Agriculture,
Nature Management and Fisheries
The Netherlands

Maritta R. von Bieberstein Koch-Weser
Director General
IUCN - The World Conservation Union

Acknowledgements

Summary

Sandra works in an Environmental Education Centre. Recently she has been getting less satisfaction from her work. She feels that some activities could be approached in a better way. What is the value of all these activities and effort? On top of that, the local council, the Centre's main source of funding, is asking more and more questions about how its money is being spent.…

Sandra would like to adopt a more systematic, slightly more structured approach to the activities of the Centre. In doing so, she comes across many different facets of evaluation.

This is the context for this detailed text on the evaluation of environmental education programmes. In Chapter 1 the book starts with the example of Sandra's centre, explains the purpose of evaluation and outlines the thirteen steps behind the evaluation process.

Chapter 2 takes a look at the broader principle of how evaluation can be introduced as a regular activity in an organisation.

Chapter 3 revisits the thirteen steps on evaluation in more detail. Here you can go deeper into the theory and practice of carrying out large scale and comprehensive evaluations of education programmes. The text is illustrated throughout with exemplars from a range of actual evaluations carried out in Holland and elsewhere.
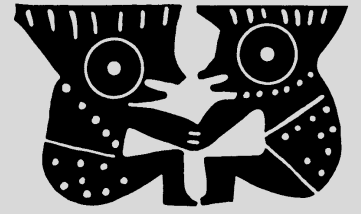
For those looking for more technical detail, the appendices go into the design of instruments and give examples for the reader to use as models.

The text addresses areas such as:

- clarifying objectives;
- identifying target groups;
- data collection methods;
- developing appropriate research tools;
- collecting, processing, analysing and interpreting data;
- compiling an evaluation report.

A list of further reading on evaluation in environmental education is included for those who want to pursue this area further.

The handbook is a useful source book for setting up evaluation for managers of non formal environmental education programmes in protected areas, environment centres, field study centres, botanical gardens, museums or historic and cultural centres.

Chapter 1

To ease you in:
evaluation step by step

The 13 steps to evaluation described in Chapter 1.

Step 1
Determining the reason
for the evaluation, or the
purpose(s) which the
evaluation is to serve

Step 2
Determining the object
of evaluation

Step 6
Listing one or more
factors which facilitate or
inhibit the achievement
of the objectives

Step 3
Defining the objectives of the product or activity

Step 4
Deciding when, and
from which groups, data
should be collected

Step 5
Selecting one or more
data collection methods

Step 7
The creation of one or more research instruments

Step 8
Deciding who to include in the investigation

Step 9
Collecting the data

Step 10
Processing and analysing the data

Step 11
Interpreting the results
and drawing
conclusions

Step 12
Writing a report on the
evaluation and the
results and conclusions

Step 13
Reaping the benefits of
the evaluation

To ease you in: evaluation step by step

In this chapter you will meet Sandra. She works in an Environmental Education Centre in the middle of the country. Sandra would like to adopt a more systematic, slightly more structured approach to the activities of the Centre. In doing so, she comes across many different facets of evaluation.

Sandra's story provides an introduction to evaluation. We use a fictional example to explain, in concrete terms, what systematic evaluation entails: what is important in evaluation, what you have to consider when setting up and carrying out evaluation, what useful purpose evaluation can serve and so on.

Although Sandra's story is fictitious, it will still provide many points of reference for people working in both large and small nature and environmental education organisations. In addition, this chapter and the book as a whole, may prove useful to people in other areas of education, for example in the field of development education.

To help you obtain a clear overview, this account follows the structure of the steps referred to in full in Chapter 3 where all the steps to implement evaluation are elaborated. For clarity, references to Chapter 3 are included in *italics*.

Sandra and the Centre

Sandra has been working three days a week at the Environmental Education Centre for five years. The Centre offers a wide range of products and services. For example: teaching packs are developed and lent out, primary and secondary schools can invite staff from the Centre to give lessons in nature and environmental education, children and adults can attend courses at the Centre, and so on.

You can see already that the Centre has several different target groups. Its work is aimed at children and adults in various settings. For example, it deals with pupils and teachers in primary and secondary schools, welfare agencies such as visitors from community centres, and ordinary members of the public. Members of the public can also attend one-day or longer courses such as a course on environmentally friendly gardening for women.

Recently Sandra has been getting less satisfaction from her work. She feels that some activities could be approached in a better way. Now and then she gets the feeling that a number of activities are producing little in the way of results. What is the value of all these activities and effort? On top of that, the local council, the Centre's main source of funding, is asking more and more questions about how its money is being spent. The Centre is increasingly being asked to demonstrate exactly what benefits are obtained from the activities. What is actually being achieved with all that money?

(Step 1: Determining the reason(s) for the evaluation, or the purpose(s) which the evaluation is to serve)

In fact, Sandra needs to carry out a systematic evaluation. With the aid of evaluation she can not only provide the council with a justification ('you see, we do achieve the results we promise'), but also keep a finger on the pulse with answers to questions like "Are we achieving our objectives?" "How can we best tailor our work to our target groups?", "What can we improve?", "Who needs to improve his/her work?"

The Organisation

First a quick look at the organisation. There are a number of people working at the Centre. There are four permanent staff. The head of the centre is Rina. She works four days in all, two at the Centre and the rest at another centre in a neighbouring town. Rina maintains contact with the funder, she is responsible for the Centre's finances and the coordination of the volunteers and trainees. Then there is Kees, who works four days a week in the Centre. Like Sandra, he is a biologist and is responsible for developing materials and for organising and running activities. The fourth permanent member of staff is Janine. She is a secretary and also works four days a week. Her job is to design and produce the materials once developed, keep a register of the materials lent out, receive visitors, etc.

The Centre also uses volunteers who help with the supervision of groups, the maintenance of the materials and the reception and care of visitors. Finally there are often students on placement at the Centre. Most are biology students. These trainees often help with the supervision of groups and in some cases will supervise a group or visit a school on their own.

To make a start, Sandra first talks to her colleagues, starting with Kees.

Consultation with co-worker Kees

Discussion between Sandra and Kees gets off to a difficult start. When Sandra says that she wants to evaluate the Centre's range of activities, Kees's reaction is that he has been doing that for a long time. He always talks to the participants and sometimes gets them to fill in a questionnaire. The participants are generally very positive when they talk to him, and even written surveys do not suggest there is anything in the range of activities, which should be changed. Sandra counters by saying that it is good that the participants are so satisfied, but such measurements of satisfaction do not indicate whether the Centre's activities have any real effect. Furthermore, she confides that the council wants to know whether the Centre is getting results, and not whether the participants are satisfied. Finally, measurements of satisfaction do not show how things should, and can, be improved and things can always be improved!

Sandra and Kees agree to evaluate one aspect of the Centre's activities. If that goes well and produces results, they will look at other aspects. The question is, which aspect?

(Step 2: Determining the object of the evaluation)
Because this is their first attempt at systematic evaluation, they do not want to devote too much time to it at this point. They decide that the object of the evaluation should be easily manageable. For this reason Sandra and Kees set out a few conditions:

- there must be some measurable effect, in other words, the activity must continue for a sufficient period;
- the target group must have the necessary skills to complete the evaluation, for example, to fill in a questionnaire;
- it must be an activity which they both do, so that they can set up the evaluation together;
- the activity to be evaluated must be run on a regular basis so that the evaluation tools which they develop can be put to regular use.

A number of activities fail to meet these conditions. For example, a course held in an afternoon produces little practical effect, let alone effects in terms of changing attitues, activities with primary school children who lack the language skills required to fill in a questionnaire and one-off activities.

Finally they settle on the nature course for parent volunteers. Does the course achieve the desired results?

The object of evaluation: the nature course for parent volunteers

The nature course for parent volunteers is intended for the parents of pupils. It is comparable to 'mothers' reading' for example. The parents assist primary school teachers in giving and organising lessons, in this case in nature and environmental education.

The course, as given by the Centre, consists of several part-days. In these part-days parents learn about plants and animals, how to use a range of equipment from magnifying glasses to saws, where to obtain materials and how to ask pupils stimulating questions.

Sandra and Kees are happy with their choice. The course is organised several times a year. This offers Sandra and Kees the opportunity to re-use the research instrument they develop, and to improve the parents' nature course using the results of the evaluation. The course is run by both Sandra and Kees and lasts long enough for the effects to be measured.

Measuring the effects

But what are they actually trying to achieve with the course? Now that they are discussing it together, they discover that they have set different priorities. Sandra mainly wants to make the parents enthusiastic about the natural world, so that they want to undertake all sorts of activities with the pupils and, in that way, stimulate the teaching staff to devote more time to nature, natural history and environmental education. Kees thinks that is important

too, but he places the emphasis mainly on what the participants need to know and be able to do. After all, these particular parents are already enthusiastic; otherwise they wouldn't be doing the course.

They decide that each of them will write down their own list of course objectives. It will make discussion easier and later it will make it easier to make a selection.

(Step 3: Defining the objectives of the activity or product)
Sandra and Kees reconvene a week later. Although they do not altogether agree about the importance of the various objectives, they have identified many of the same ones. Both have formulated objectives concerning what the participants should know and what skills they should have on completion of the course. They have also tried to quantify the kind of attitude the course is trying to engender. This last point proves really difficult. An objective such as 'the participants feel involved in nature' is, after all, fairly vague.

Eventually they select the following objectives as a starting point.
The participants:

    a. can recognise and name common plants and animals in their surroundings and know what they require from their habitat;
    b. can recognise evidence and give examples of food chains;
    c. know how to use common equipment for investigation including magnifying glasses, pruning shears and saws and can explain to pupils how they should be used;
    d. know where to obtain materials and are creative in collecting materials which may be useful for nature studies;
    e. are capable of asking pupils questions which encourage self-learning through discovery;
    f. in their contacts with pupils, place the emphasis on being involved in, enthusiastic about and active in the natural world, rather than on imparting factual information;
    g. have a better understanding of how caring for the environment in and around the school can be put into practice;
    h. feel involved in and responsible for the natural environment and are enthusiastic about working in and around the school.

Setting up the evaluation: when, how and from whom should data be collected?

The objectives referred to above are formulated at the level of the participants, in other words, the parents. However, the ultimate goal is that the *pupils* acquire knowledge and skills and that they are interested in the natural world. The teaching staff must work towards this with the assistance of the parents. Sandra wonders from whom she should collect data.

(Step 4: Deciding from which groups data should be collected)
There are four options:
- the pupils: what do the pupils know and what can they do; what have the parents contributed to this, either directly or via the teachers?
- the teaching staff: do the teachers feel stimulated and supported by the parents?
- the parents on the course: what are the effects of the course on the participants?
- a combination of two or three of the above.

Kees proposes limiting the evaluation in the first instance to the parents. "We need to know what the parents have got out of it before we can measure any effects on the teachers and/or pupils."

So what do the participants get out of the course? According to Kees this must be translated into the knowledge, attitude and skills they have after the course, compared with the knowledge, attitude and skills they had prior to the course.

(Step 4 continued: Deciding when data should be collected)
Sandra agrees. "But actually, you are now saying two things." she says. "First we have to collect data before and then again after the course. If we only collect data after the course, we will not know how much we have built on the knowledge, skills etc. the parents already had. Then we obviously can't say what effect the course has had!"
(See also Step 6: Listing one or more factors that facilitate or inhibit the achievement of the objectives.)

The instruments
Now they must decide which instrument or instruments to use.

(Step 5: Selecting one or more data collection methods)
"What are the options?" For Kees, the options are interviews, a written questionnaire or observation of the participants. Kees considers interviews have the advantage that they can probe quite deeply into the way the participants think about the natural world and the environment and about the way this is passed on to children. He also feels it is only by observation that you can really discover what skills the participants have. However, he notes that observing and interviewing are both extremely time-consuming, particularly since they are going to take measurements before and after the course. "We don't have that sort of time!" he says.

They opt for a written questionnaire, but they are concerned about whether they can convert all the objectives properly and reliably into written questions. At least they will try. If it doesn't work out, they can discuss it with Rina and see if she has any other suggestions. Perhaps one of the trainees could then conduct interviews and process the results.

The questionnaire

For Sandra and Kees the next step is to consider the facilitating and inhibiting factors and write the questionnaire. They are going to 'operationalise', i.e., convert the objectives into a concrete series of questions and possible responses.

(Step 7: The creation of one or more research instruments)
Sandra and Kees do not want to take up any more of the participants' time than necessary. This means that the questionnaire should be as short as possible, that the questions must be clearly stated and that the participants should not have to formulate their own answers, but simply select an option from a list of possibilities.

Sandra and Kees start with their list of facilitating and inhibiting factors. They feel that various factors influence the impact of the course *(see Step 6)*, such as:

- the gender of the participants (for instance this might influence skills in using tools and their inclination to do so);
- the initial status of the participants (what skills and knowledge they already possess);
- their reasons for participating (do they really want to do something for the school, or are they attending more for their own benefit?);
- satisfaction with the design and running of the course and their reception at the Centre.

Operationalisation of the influential factors

Of course, it is simple enough to establish the gender of participants. The question looks like this:

1. Are you

    0   male?
    0   female?

For now, they skip over the initial status of the participants. This will be dealt with later, when they come to put the objectives into operation. They attempt to chart the reasons for participation in the course by means of a number of statements. In the first five statements, the participant is central. The participant takes part in the course largely for his/her own benefit. In the last four statements, the focus is on the pupils and teachers.

2. How much did the following considerations influence your decision to take part in the course?

    1 = not at all
    2 = only slightly
    3 = quite a lot
    4 = very much

I like meeting new people	1	2	3	4
I want to find out more about nature and the environment	1	2	3	4
I want to improve my gardening skills	1	2	3	4
I want to make better use of my free time	1	2	3	4
I want to have some status at the school my children attend	1	2	3	4

I want to encourage children's involvement with the
   natural world in different ways   1 2 3 4

I want to help children to have a positive experience
   of the natural world   1 2 3 4

I want to help children to appreciate the natural world
   more fully   1 2 3 4

I want to be helpful to the teaching staff, to enable
them to devote more attention to the natural world
   and the environment   1 2 3 4

Finally they come to satisfaction with the design and running of the course and reception at the Centre. Here, Sandra and Kees distinguish between a number of aspects, which they make operational as follows:

3. Please rank each of the following statements by circling the number which best corresponds to your view.

   1 = totally disagree
   2 = largely disagree
   3 = mildly disagree
   4 = mildly agree
   5 = largely agree
   6 = totally agree

The course:
I thought the course was well designed   1 2 3 4 5 6
I thought the content of the course was good   1 2 3 4 5 6
I was happy with the amount of time spent on the course   1 2 3 4 5 6
I thought the course elements were suitable for
   primary education   1 2 3 4 5 6
I found the course inspiring   1 2 3 4 5 6

The course instructor:
I found that the instructor had a good knowledge
   of the subject matter   1 2 3 4 5 6
I was happy with the teaching skills of the instructor   1 2 3 4 5 6
The instructor showed a great deal of enthusiasm   1 2 3 4 5 6
The instructor organised the course well   1 2 3 4 5 6
The guidance provided by the instructor was good   1 2 3 4 5 6

The Centre:
The reception at the centre was good   1 2 3 4 5 6
The tea and coffee were satisfactory   1 2 3 4 5 6
The course information provided by the centre was good   1 2 3 4 5 6

Sandra and Kees decide to include questions 1 and 2 in the survey given to the participants prior to the course, i.e. in the initial survey. Question 3 is to be included in the final survey.

Making the objectives operational

Sandra and Kees think they can make a reasonable job of the influential factors. However, it is not so easy to make the objectives operational. Objectives relating to knowledge should be fairly straightforward, but those relating to skills with equipment, dealing with pupils and feelings involved with the natural world are not so easy to formulate.

Sandra and Kees split the work between them. Kees will take responsibility for the food chains and equipment skills, while Sandra will try to work out a formula for the relationship with pupils and the relationship with the natural environment. Afterwards they comment on each other's questions. They agree beforehand not only to look at the operation, but also at the clarity of the questions.

First the food chains. Kees has thought up the following questions:

a. Complete the following food chain

.............…….… - ....….… - human beings

b. Monique catches a pike in a lake. She eats the pike. The lake also contains smaller fish which eat water fleas. There are also algae growing in the lake. What food chain can be traced from the story above?

0   Monique - pike - water fleas - smaller fish - algae
0   pike - smaller fish - water fleas - algae
0   water fleas - smaller fish - pike - Monique
0   algae - water fleas - smaller fish - pike - Monique

c. What is the order of the food chain formed by fallen leaves, herons, moles and earthworms?

0   earthworms - fallen leaves - moles - herons
0   fallen leaves - earthworms - moles - herons
0   herons - moles - earthworms - fallen leaves
0   moles - herons - earthworms - fallen leaves

For skills with common pieces of equipment, Kees would like to evaluate through observation because that is the way to find out how well people use them. However, since they have opted for a questionnaire, he can see no way of charting these skills other than to ask a question which measures how the participants perceive their own skills.

d. This section is about your experience with common pieces of equipment. Please indicate whether you have ever used the equipment concerned, and if so, how well you got on.

		have done:	I got on:	
	never done	badly	reasonably	well
working with a saw outdoors	0	0	0	0
working with pruning tools	0	0	0	0
using a telescope / binoculars	0	0	0	0
using a magnifying glass	0	0	0	0
using a soil borer	0	0	0	0
using a 'search list'	0	0	0	0

Sandra decided to investigate the relationship with pupils with a number of statements. The participants have to indicate how much they agree with the statements.

e. Please read the following statements then indicate whether you agree or disagree by circling the number which best corresponds to your view.

1 = totally disagree
2 = largely disagree
3 = mildly disagree
4 = mildly agree
5 = largely agree
6 = totally agree

The transfer of knowledge is the most important aspect of nature and environmental education	1 2 3 4 5 6
It is important in nature and environmental education for pupils to do experiments themselves	1 2 3 4 5 6
Nature education is about knowledge of plants and animals	1 2 3 4 5 6
The main objective of nature education is that pupils learn to appreciate the natural environment	1 2 3 4 5 6
It is absolutely essential for good nature education that pupils deal with concrete materials and living organisms	1 2 3 4 5 6
If children are enthusiastic about the natural environment then you have achieved a great deal	1 2 3 4 5 6

Finally the question relating to whether the participants feel involved and responsible. Sandra wants to investigate this by asking the participants about their behaviour.

f. Please indicate below how often you do the things listed

1 = never
2 = sometimes
3 = often
4 = very often

Talk to other people about how they can contribute to a cleaner environment	1	2	3	4
Try to persuade other people to behave in a way that is better for the environment	1	2	3	4
Comment when you see people behaving in an environmentally-unfriendly way	1	2	3	4
Read natural history and/or environmental publications published by nature and environmental organisations	1	2	3	4
Read newspaper reports on nature and environmental subjects	1	2	3	4
Watch or listen to nature and environment programmes on the TV and radio	1	2	3	4

Questions a) to f) are to be included both in the pre-course and post-course measurements. This way Sandra and Kees can determine whether there is any change in the knowledge, skills, attitude and behaviour of the participants. These questions can also provide insight into the status of the participants at the start. Perhaps they can distinguish different groups based on the initial measurements, such as participants with a lot or a little prior knowledge, environmentally friendly and unfriendly participants, and the like.

Administration of the questionnaire

Sandra and Kees have reached the stage where they want to give the questionnaire to an initial group of participants. They have worked out carefully which questions should be asked before and which after the course and have asked Janine to do the layout.

(Step 8: Deciding whom to include in the investigation)
"What shall we do," says Kees, "bother the whole group or just a few participants?" "Well," Sandra replies, "there are only eighteen people in the group, so it wouldn't be too hard to process questionnaires from all of them. Also, I think it would seem strange to them if we asked some and not others. And then how would we know who to select? We don't know anything about them."

(Step 9: Collecting the data)
On the first day of the course the participants are given a cup of coffee and a short introduction and then they are asked to fill in the questionnaire. It takes them nearly an hour to complete it. "That's a long time! We must see about making it shorter," thinks Sandra.

When the participants hand in the initial questionnaires, these are numbered. By using the same number for the same participant in the second measurement, Sandra and Kees will be able to link the answers to the two

questionnaires when they are processing them. Then they can determine whether the course has had any effect on a given participant.

There is some discussion between Sandra and Kees about how to set up the post-course survey. Initially they thought they would ask people to fill in the final questionnaire on the last day of the course. This was so that they could be sure to get responses from all the participants and would not need to make follow-up phone calls. However, since the forms take so long to complete, they think it would be a waste of course time. Finally they decide to hand out the questionnaire and a return envelope to the participants, who will then have two weeks to complete them and send them back. After two weeks, if too few (less than 14) have been received, Janine will make a few phone calls. "So we will need to know who has and who hasn't returned their questionnaire," says Sandra, "otherwise Janine won't know who to phone." "We can work that out from the numbering," says Kees. "We just put the numbers on the questionnaires."

Processing and interpreting the information

Eventually Sandra and Kees get 15 questionnaires back. They are happy with this response. Now to the processing!

(Step 10: Processing and analysing the data; Step 11: Interpreting the results and drawing conclusions)
The Centre does not have computer software for analysing data, so Sandra and Kees make do with a calculator. However this limits what they can do. If they decide to proceed with evaluation, they will have to talk to Rina about getting a software package.

So what will they calculate, and how? Here are some examples.
First Sandra and Kees check whether the participants have learned anything - questions a), b) and c). Each participant is given a knowledge rating. For a correct answer to question a) they receive two points; for a correct answer to b) or c) they receive one point.

Next they calculate the average of all the participants. The average score at the start of the course was 3.6, and at the end it is 4. In the final measurement all the participants answered the questions correctly. That does not mean the participants have learned a lot because the score was high at the start.

Next they look at handling equipment - question d). Here there are greater differences between the participants at the outset. Before the course a number of participants had never used a saw, a soil borer or a search list before. At the end of the course they all indicate they have used all of them: in other words, everyone has ticked either 'badly', 'reasonably' or 'well'. In the post-course survey they all ticked 'well' except in the case of the pruning tools. Answers relating to pruning tools vary between 'badly', 'reasonably' or 'well'. Therefore more attention should be paid to the use of pruning tools.

Getting on with pupils - question e): Sandra and Kees decide to calculate the mean of all participants for each statement. First they convert the answers into numbers:

    1 = totally disagree
    2 = largely disagree
    3 = mildly disagree
    4 = mildly agree
    5 = largely agree
    6 = totally agree

The mean values for the initial and final measurements are as follows:

	Before the course	After the course
The transfer of knowledge is the most important aspect of nature and environmental education	4.3	2.4
It is important in nature and environmental education for pupils to do experiments themselves	4.5	5.5
Nature education is about knowledge of plants and animals	4	5.5
The main objective of nature education is that pupils learn to appreciate the natural environment	5	5.8
It is absolutely essential for good nature education that pupils deal with concrete materials and living organisms	4.9	5.7
If children are enthusiastic about the natural environment then you have achieved a great deal	3.8	5.5

Sandra and Kees are very pleased with these scores. The post-course scores are closer to what they consider desirable. The scores for statements like 'The transfer of knowledge is the most important aspect of nature and environmental education' have dropped, while those for statements like 'The main objective of nature education is that pupils learn to appreciate the natural environment' have risen.

In terms of the sense of involvement and responsibility of the participants - question f) - again Sandra and Kees calculate the average for each statement. They are less pleased with these scores, as there is scarcely any difference between the initial and final scores. But then, should you expect an improvement in these scores after only one course?

Satisfaction of participants with the course - question 3) - here the scores for all the items except two are higher than 4. This is very good, since it means that the participants are reasonably satisfied.

The two exceptions are 'design of the course' and 'time spent on the course'. Opinions about the design of the course vary widely, fluctuating between 1 and 5. Time spent on the course produces a range of scores, too. But do they think the course is too long or too short? Sandra and Kees regret not having asked this. They resolve to do so in any future measurement, so they can find out what the participants thought was wrong.

Finally Sandra and Kees look at the motives and objectives of the participants - question 2. Here they try to put the participants in two categories:

Group 1: score high on own interests and low on interests of pupils and teachers;
Group 2: vice versa.
This categorisation seems to work well, although there are some people who get a high score in both groups of items.

Sandra and Kees pick out just the participants who clearly fall into one or other group: 12 in total. Five fall into the first and seven into the second category. Next they refer back to some of the questions for those participants.

This shows:

Knowledge: All group 2 participants answered all the questions correctly at the beginning. This does not apply to group 1 participants.

Skills: The group 2 participants prove to have learned more than the group 1 participants. In other words, the difference in scores before and after the course was greater for group 2 than for group 1. Were group 2 people more open to learning, or was it pure coincidence? "Actually you need a computer programme to determine if there is a significant difference, or so I've heard," says Kees.

Getting on with pupils: the answers of group 2 participants varied more widely (more 1, 2, 5 and 6) than those of group 1 participants (more 3 and 4). In the pre-course measurement the scores were closer together. From this Sandra and Kees conclude that group 2 people are more inclined to form their own opinion. Group 2 has done more with the different facets of the course than group 1. Or could this also be a coincidence?

Reporting and drawing conclusions

Sandra and Kees are very pleased with the results they have obtained. They decide to put it all down in writing.

(Step 12: Writing a report on the evaluation, and the results and conclusions)
The fact that Sandra and Kees are writing up the evaluation systematically presents various opportunities. First, it will help with the organisation of a subsequent evaluation. Secondly they will be able to disseminate the report. The material can be disseminated both internally for discussion with Rina and externally, for example to the council.

(Step 13: Reaping the benefits of the evaluation)
The report also presents the conclusions they reached about the way the course is set up and run. First, they want to devote less time on the course to food chains, since the people on the course were already well enough informed. Instead they want to spend more time on practical skills such as the use of pruning tools, how to prune and what to look out for whilst pruning. Finally, when they are advertising the next course, they want to place more emphasis on the fact that the parents' nature course is primarily intended to provide support for teachers and pupils. Of course a little self-interest is not out of place, but it should not be emphasised.

They will also use the findings from this evaluation for a subsequent course. In that case they definitely want:

- to make the questionnaire shorter;
- to ask for an explanation of negative responses to the points about satisfaction;
- to make operational involvement and responsibility in a different way.

Chapter 2

Introducing evaluation as a regular activity in an organisation

Introducing evaluation as a regular activity in an organisation
What do you need and how do you approach it?

In this chapter we discuss the following subjects:

- as an organisation, what do you have to evaluate and why should you do it?
- what conditions must be met to enable you to implement regular evaluation?
- how can you conduct a trial run so that you know what to expect?
- how can systematic evaluation be brought into the organisation?

As an organisation, what do you have to evaluate and why should you do it?

Evaluation means making a critical examination of what something is worth. What the 'something' may be has already been examined in Chapter 1. The main objects of evaluation are likely to be:

- the functioning of all or parts of the organisation;
- the 'market position' of the organisation compared with other providers;
- the quality of and/or interest in the current range of provision as a whole;
- one or more products for particular client target groups;
- one or more activities for particular participant target groups.

The question of why such things should be evaluated has already been discussed in Chapter 1. In brief, the possible purposes of evaluation include:

a. reporting to a funding body (accountability);
b. to be able to provide participants and target groups with information (PR function);
c. to monitor quality (keeping a finger on the pulse);
d. to improve quality (learning from experience).

Reasons for carrying out more, or more systematic, evaluation as an organisation relate to combinations of the above. For example, it might be to determine whether an activity corresponds sufficiently to the interests of the target group, to determine what participants learn from it, to discover what aspects of a product can be improved, to make the value of your work clear to administrators, clients and others and so on. In short, as an organisation you only carry out evaluation if it is in your interests to do so.

A survey of educational establishments a few years ago showed that many appreciate the value of evaluation and are implementing it but that there are three recurrent problems:

- lack of time;
- lack of funds;
- lack of expertise.

In addition there are establishments which question the purpose of evaluation. This may simply be due to a lack of knowledge and experience.

However, it can also be due to unsuccessful attempts or a negative experience, or the difficulty of striking a proper balance between what evaluation can deliver and what it costs.

These problems can only be solved by targeted investment. However, the situation will vary according to the size of the establishment. A large organisation may decide to free up time, funds and personnel for, say, a year, in order to gain insight into what evaluation can mean for the establishment and what it would cost to introduce it. A small organisation would often be hard put to make such an investment. However, in that case there is probably scope for obtaining a special temporary budget allocation. Educational organisations which, for example, are part of a local authority, can attempt to make use of budgets for professional development, quality control, improving market orientation, efficiency or effectiveness or whatever 'label' may suit the purpose.

Organisations which receive support from, say, a government department, can try to put forward a case for a one-off investment.

A proposal stands more chance if it is described in a short but clear plan, setting out the objectives, activities, costs, benefits and planning. One possible benefit is an operational plan for evaluation within one's own organisation. Another is a description of the steps followed to reach this point because it will probably be of interest to other parts of the authority or to other institutions associated with the department.

Both small and large organisations need to get to grips with, and make decisions about what can and should be evaluated. This then needs to be translated into a small project for a limited period. The project should enable you to answer the following questions:

- What are the main reasons for the organisation to consider introducing more systematic evaluation? What aspects of the organisation, the work or the services offered, does it seem particularly appropriate to evaluate?
- What conditions must you be able to meet in an organisation in order to carry out such an evaluation and obtain results? To what extent does the organisation meet these conditions already and on which points can the organisation meet the conditions in future and what needs to be done to effect this?
- Assuming that you have decided to implement on-going evaluation, how can this be approached? (It is most effective if a pilot study is used to create a plan of action, since that is the best aid to concrete decision-making.)

The first question is addressed elsewhere. Here we will restrict ourselves to answering the second and third questions.

What conditions must be met in order for an organisation to introduce regular evaluation?

If you wish to tackle evaluation seriously, you must meet certain conditions. If certain conditions are not being satisfactorily met, it is advisable to address this problem first of all. Furthermore, in considering all the conditions, it is not a matter of whether they are met in full, but whether they are fulfilled to the necessary extent. There is always room for improvement. It is up to the organisation to decide what is 'sufficient' at any given point, given the time and means available, and given the purpose the evaluation is to serve. If the aim is to produce an official report or if an activity which is strategically important to the organisation is due for a radical overhaul, you will be inclined to set more stringent requirements and to make a greater investment.

In short, it is not necessary to meet all the conditions in order to make a useful start on evaluation.

The list can be used preferably with the team, to determine the extent to which you would like to fulfil certain conditions better. At first sight it may seem that it will require considerable time and effort. In practice, however, many of the conditions relate to 'having given a good deal of thought to your work'. And that is very important in itself, irrespective of evaluation.

The conditions relate to the following four categories:

- products and activities which are suitable for evaluation;
- clear division of responsibilities and cooperation and openness in the team;
- time, material, space;
- expertise.

Products and activities which are suitable for evaluation

The product in general

1. Activities and products of sufficient importance
Organisations which carry out many fragmented activities and/or have only fleeting contact with the participants or which restrict themselves, say, to the distribution of short pamphlets, must question whether it is worthwhile to carry out systematic evaluation. That is, whether it is worthwhile to do any more than run through a simple checklist from time to time or ask members of target groups, clients or participants very briefly about their satisfaction and requirements.

Products such as teaching materials or a course can be evaluated with relative ease by experts and users. If you wish to determine the extent to which target groups learn from a product, the evaluation begins to resemble that of an

activity, since one is then evaluating the use of a product within the framework of an activity. For this reason we will refer only to 'activities' below.

The activity in general

2. Clear definition of the activity
It must be possible to describe an activity clearly. It must be clear where and when the activity takes place.

The activity should be reasonably stable over time, in other words not always being set up and implemented differently. Evaluation of a one-off activity is possible, but you would not invest so much in it.

It is also important to be able to assume that the activity will be implemented as planned or that you can determine the extent to which it was or was not done. Also, when evaluating the effects of the use of a product, it is important to determine the extent to which the product was actually used.

The objectives of the activity

3. Clear objectives
An activity is always evaluated against certain criteria. Often the most important criterion is whether the activity has any effect. The intended effects are referred to as objectives and the evaluation determines if the objectives were achieved. That can only happen if these objectives have been clearly formulated.

4. Measurable results
Objectives must not only be clear, they must be measurable. For example: 'to contribute to the development of a sustainable society' is not an easily measured objective.

The implementation of the evaluation

5. Availability of suitable times
The scheduling of the activity must make it possible to collect data prior to, during and/or after the activity.

6. Accessibility of participants
Where data must be collected before or after an activity, you have to know how to contact the participants.

7. Willingness of participants to cooperate
Participants must be willing to help with providing information or to assist in collecting data from others , for example asking teachers to collect data from their pupils.

8. Language skills of participants
All methods of data collection, apart from observation and assessment by third parties, require some language skills on the part of the participants.

Interpreting and using the results

9. Expectation of learning results
You must have some expectation that the participants will achieve certain learning results. Without such an expectation there is no point in evaluation and you probably would not even embark on the activity. However, this is not entirely self-evident. Sometimes participants already have the skills and knowledge at which the activity is aimed. Sometimes the majority of people who sign up for an activity are already very active and come more for the social contact than to learn anything new. Sometimes participants are not very open to the planned learning processes due to lack of interest or motivation, for example, when participation is compulsory.

10. Assessment of the learning results
The significance of the results is not always obvious. You must be able to make a comparison with something else. For example, are they roughly what you could have expected? If an activity has not previously been systematically evaluated, it can be difficult to determine whether you should regard the results as good or poor. Only by carrying out several such evaluations or by using experience gained elsewhere can you obtain comparative material.

11. Clear relationships between cause and effect
As long as you are only evaluating to determine to what extent objectives are achieved and to present a justification based on that, the measurement of the results is sufficient. The only complication is that you must be able to establish that the results really are the effect of the activity and not of something else.

There are various ways of making a reasonable case to say that the results could not have been caused by something else. For example, it is plausible that the results were not obtained elsewhere:

- if the activity takes place in a relatively short, continuous period then the participants will have little chance to learn similar things in the meantime;
- if the activity is aimed at fairly specific goals then the knowledge and skills which cannot easily be picked up elsewhere;
- if clear links can be established between the means used (content, working methods, materials, activities) and the objectives.

It is even more useful if you have some idea of how the involvement of the participants in the activities can lead to the desired learning results. In other words, how the participants can acquire the envisaged knowledge, skills and attitudes. We are now in the realms of educational psychology.

Clear division of responsibilities and good cooperation

12. Clear implementation structure
Sometimes evaluation does not provide the information you need. To prevent this it is useful if it is clear from the outset who is responsible for what.

13. Cooperation and openness in the team
Evaluation of your own activities makes most people feel anxious. You must have mutual trust in the team so that you can determine openly and together where improvements are necessary and who should be responsible for taking action.

Time, materials, space

14. Working time
As with many things, there is no point in starting on a course of systematic evaluation if you are not prepared to invest the necessary time.

A distinction can be made here between the investment of time required for setting up an initial evaluation and the time required to implement regular evaluation using a previously developed set-up.

In the long term it is sensible to allocate an average of 5% of the total available time to evaluation. Of course, you would only invest that much time if the results were going to be put to good use - otherwise you might as well not start at all.

15. Hardware and software
In Chapter 1 we discussed whether the processing of the information could be carried out manually and under what circumstances one would really need to use a computer. Without a computer you will probably restrict yourself to keeping a simple tally, working out percentages and means and investigating differences between groups or opinions. Such differences can only be described; they cannot be tested for statistical significance. In that case you don't know whether or by how much a difference really means something, or whether you can attach any significance to it. Testing for statistical significance is easiest if you are using a computer.

The same applies to analysing the reliability of a scale (a number of questions which measure the same concept); if the answers to those questions are sufficiently consistent, the scale is said to be reliable.

There are various computer programmes available for data analysis and you may need some expert help when making your choice.

16. Space
Physical space is required for evaluation work. For example space will be needed for team consultations, for compiling a questionnaire, for storing returned questionnaires.

17. Expertise
The specific knowledge and skills required to carry out your own evaluations are described elsewhere in this publication. In practice they largely fall into the following three categories:

- setting up, organising and supervising the evaluation process as a whole;
- developing a suitable research instrument;
- processing and analysing the data obtained and drawing conclusions.

In an organisation of any size one can opt to have one or two members of staff for each category. If there are at least two expert members of staff per category, this spreads the load and the risk.

In a small organisation it is more likely that one or two members of staff will work on the entire process.

Also, something like evaluation is best learned by trying it out in practice. Small scale projects enable you to test evaluation methods without investing too much time. It might be a pilot project for a larger evaluation to follow.

How can systematic evaluation be introduced into an organisation?

Evaluation can be introduced into an organisation by following the steps described below.

Initially, a year is set aside for the staff to familiarise themselves with evaluation. During this period:

    a. time is set aside;
    b. duties are allocated;
    b. if necessary, hardware and software are obtained;
    c. expertise is developed where required;
    d. some initial evaluations are carried out;
    e. experiences are discussed.

Subsequently a period of time - say two years - is set aside to build up experience. During this period:

    • an evaluation plan is developed for the period;
    • the planned evaluations are carried out;
    • the results are put to use;
    • experiences are discussed.

It is advisable to repeat the evaluation of a particular activity a number of times, as this makes it possible to determine the degree of variation in results between groups of participants. It also indicates how much time is required for regular evaluation of the relevant activity.

The evaluation of a number of new activities is also recommended, since this provides further experience with Steps 3 to 7 and provides greater insight into the results of the different types of activity offered by the organisation.

To round off the introductory phase the following steps may be taken:

    • introduction of evaluation into multi-year planning, the plan of
      action and the budget of the organisation;
    • incorporation of the various evaluation duties into the job
      descriptions of staff;
    • incorporation of the various activities into organisational processes
      (routines, agreements, etc);
    • incorporation of checks and balances to monitor the implementation,
      quality and usefulness of evaluation.
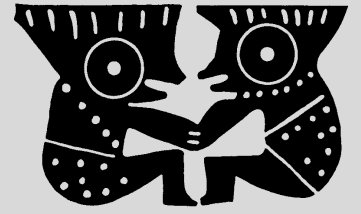
Chapter 3

Thirteen steps
to evaluation

This chapter sets out in more detail the 13 steps to evaluation described in Chapter 1.

Step 1
Determining the reason for the evaluation, or the purpose(s) which the evaluation is to serve

Step 2
Determining the object of evaluation

Step 6
Listing one or more factors which facilitate or inhibit the achievement of the objectives

Step 3
Defining the objectives of the product or activity

Step 4
Deciding when, and from which groups, data should be collected

Step 5
Selecting one or more data collection methods

Step 7
The creation of one or more research instruments

Step 8
Deciding who to include in the investigation

Step 9
Collecting the data

Step 10
Processing and analysing the data

Step 11
Interpreting the results and drawing conclusions

Step 12
Writing a report on the evaluation and the results and conclusions

Step 13
Reaping the benefits of the evaluation

Thirteen steps to evaluation

Step 1: Determining the reason for the evaluation or purpose which the evaluation is to serve

As we have seen, it is important to take time to consider why an evaluation is carried out. Possible purposes of evaluation include:

  a. to report to a funder (accountability);
  b. to provide participants and target groups with information (Public Relations);
  c. to monitor quality ('finger on the pulse');
  d. to improve quality (learn from experience);

(a) reporting to a funder
The point here is to be able to show the results of your work. You must be able to demonstrate that you have achieved something with the funds you receive. This can best be done by showing whether the results promised or agreed beforehand have actually been achieved.

The funding body must understand the way in which the evaluation has been carried out. This will allow them to determine whether the results which you are reporting were obtained in an objective and reliable way and whether they are really representative.

(b) providing participants and target groups with information
The positive results from evaluation can help in promoting interest and confidence in your products and activities.

(c) monitoring quality
Evaluation is an instrument for investigating what exactly you are doing in your work. You can use evaluation to find out whether you are meeting your set objectives, if you are adopting the right principles and whether your work meets the pre-set quality criteria.

(d) improving quality
Evaluation can give you pointers to possible improvements in your work, since you will then have a clear idea of the reasons for the results. What part of your work is particularly effective, what is less so and so on?

It is important to set out clearly what purpose or purposes the evaluation must serve.

Step 2: Determining the object of evaluation

Once the purposes of the evaluation are established you must define what exactly you are going to evaluate.

Possible objects of evaluation include:

- the functioning of all or parts of an organisation;
- the 'market position' of the organisation;
- the quality of, and/or interest in, the current range of products and services as a whole;
- one or more products for particular target groups (tangible goods which can be distributed to 'clients');
- one or more activities for particular target groups.

Selecting the object of evaluation can be particularly difficult if you are working with or for several target groups.

2.1 Working with several target groups

By way of example, we will look at education for pupils and/or teachers. Four possible scenarios are:

- education for pupils;
- pupils and teachers are regarded as a single system;
- pupils remain the ultimate target group, but it is the teachers who work with the pupils, and the organisation works only with the teachers;
- the teachers are the target group, and it is regarded as their responsibility to provide an effective educational programme for the pupils.

An example: assignments for field work by pupils can be used during the field work activities organised by the environmental education organisation, but can also be distributed to teachers, for use in their activities with pupils. Some examples are given below:

	products	activities
teachers	course material handbook on the organisation of fieldwork	course for teachers organisation of fieldwork with pupils
pupils	field work assignments	field work activities

The situation of the dual target group means that there is more to evaluate and a choice might have to be made between evaluating one or the other.

Step 3: Defining the objectives of the product or activity

Here, we distinguish between the evaluation of a product and the evaluation of an activity.

Evaluation of a product

If an organisation develops a product it has certain objectives in doing so. Generally speaking these would be:

- that the product is positively received and appreciated;
- that the product is taken up;
- that the product is used;
- that the product is effective at achieving the desired results.

Evaluation of an activity

The objectives of educational activities might be categorised as follows:

- process objectives: the set-up and content to be achieved, the number of participants, the quality of the implementation; satisfaction among those involved;
- effect objectives: the knowledge, skill, training or affective and behavioural objectives that you hope to achieve with the target group.

Establishing and formulating the objectives can best be done in a number of steps. In practice it is simply not possible to arrive at a good useable objectives in one fell swoop. Systematically following a number of steps improves the chances of success.

Section 3.2 provides clarification and assistance in distinguishing between different kinds of objectives and section 3.3 does the same with regard to the method of formulating effect objectives. Section 3.4 describes a number of steps which can be used to formulate an objective effectively. There should be sufficient material in these three sections to enable you to make an effective evaluation of an educational activity.

Sections 3.5 and 3.6 supplement and elaborate on section 3.2 with regard to distinguishing objectives in a multiple target group. They also give further specifications on effect objectives, criteria and standards.

3.1 Evaluation with regard to the four objectives of an educational product

An organisation has certain objectives when developing a product. Generally these are (1) that the product is positively received, (2) that it is taken up, (3) that it is used, and (4) that it is effective.

Let us look more closely at these four aspects:

(1) Various groups can determine the **reception** or appreciation of a product: experts, potential clients, people who have received or purchased the product and people who have used the product.

A product can be assessed on a number of features:

- Those that say something about the quality of the product. This relates to issues such as whether there are any factual inaccuracies in the product or whether any important elements are missing. For example, teaching material about sustainability must not say that the use of earthenware mugs is more sustainable than the use of plastic beakers, *if that is not so, or is not known to be so*.
- Those that are promising in terms of demand for the product and its use and effectiveness.

We can use the information available in relevant literature concerning factors which contribute to these features. Information about take-up and use is gathered in marketing and innovation surveys. Information about effectiveness is often based on research into teaching methods and educational psychology.

(2) The **actual take-up** of a product is also a valid evaluation criterion. However, possession of a product does not necessarily mean that you familiarise yourself with it so, 'familiarisation' is a further criterion. A next step is finding out if the product is adopted for use. To summarise, the essential steps are interest, take-up or purchase, familiarisation and adoption.

(3) A further criterion is the actual **use** of the product. This coincides with the concept of 'implementation'. After all, the use of one or more products is usually part of the implementation of an educational programme. The degree of implementation of the activity then relates to making greater or lesser use of those products.

Instead of saying that a product is or is not used, it is more correct to say that a product is used to a certain extent and in a certain way. It is for this reason that we often refer to the 'degree of implementation'.

The use of a product can also be measured. This can be done in various ways: you can observe the activity being undertaken, ask the course leader to keep a log, ask the participants after the course to check off what was done or what subjects were covered and so on.

(4) Finally, the **effectiveness** of a product can be investigated. This is generally based on the use of the product within the framework of an educational activity. In other words, studies of the effects of using the product are incorporated into the evaluation of that activity.

3.2 Distinguishing types of objectives for an activity

The objectives of an educational product or activity can relate to obtaining knowledge and insight, learning skills, developing values, standards, attitudes and

behaviour. All such objectives concern the changes or effects, which one hopes to bring about in the participants. Thus we refer to these as effect objectives.

Aims, which relate to the development of values, standards and attitudes, are also sometimes called training objectives or affective objectives.

Aims, which relate to changes in behaviour, are not always welcomed - they can be perceived as social engineering. There is a discussion of this subject in the box below.

To achieve effect objectives, the activity concerned will have to meet certain conditions. We call these **process objectives** that include aspects like good organisational set-up, adequate content, smooth running, etc.

In addition, there are process objectives that are not directly related to the desired effects, but more to conditions to be met on the part of the organisation. For example, an activity must attract a sufficient number of participants and be carried out efficiently.

Many providers of educational activities will not only hope for effects in their participants, but will also hope to contribute to certain desirable effects in society as a whole. Such 'social effects' are, however, difficult to evaluate, so we will restrict ourselves here to the evaluation of the effects in the participants.

Some notes on the formulation of effect objectives and of objectives in terms of behaviour

Some people feel that it is not desirable to set behavioural objectives. Arguments to this effect include:

- changes in behaviour should not be the job of the school;
- in pedagogical terms it is undesirable to seek to change behaviour;
- it is not possible to exert the necessary influence to change behaviour because behaviour depends on environmental factors as well as the results of learning;
- it would be difficult to set changes in behaviour as a concrete objective because our understanding of what contributes to sustainability is continually changing.

However, many schools understand the importance of more environmentally-friendly behaviour and take up the challenge, partly 'to do something about nature and environmental education'. They find it entirely natural to direct their efforts towards the behaviour of the pupils while allowing the pupils to retain some freedom to determine their own course.

Another point of view is that setting objectives at the level of individual behaviour places responsibility for nature and the environmental problems too one-sidedly on the individual, disregarding the social mechanisms, which threaten the quality of nature and the environment.

3.3 Guidelines for the formulation of effect objectives

The following guidelines apply to formulating objectives.

Do not refer only to the subject matter, but also indicate the behavioural aspect
What the pupil must be able to do in relation to that subject matter. Learning itself is not an observable feature in the pupil, but a process of which the results can be measured For example:
Incorrect: Learn the parts of a plant.
Better: When shown a picture of a plant, the participants must be able to indicate and name its different parts.

Do not describe the activities that the teacher may propose to undertake
A teacher gives instruction, sets tests, marks work, holds discussions etc. All these activities are steps in the direction of the objective, not a goal in themselves. It may be useful to describe them but then it is better to refer to them as a 'teaching programme' or some such, and not as 'objectives'. This prevents concepts from becoming confused, while it makes it clear that the same objective may sometimes also be pursued using a different programme.
Incorrect: The course tutor shows how an experiment is done.
Better: After watching an experiment, the participant can reproduce the experiment without errors.

Do not describe the activities required of the pupils during lessons
Learning experiences such as seeing, participating, learning are not the same as learning results. The same experience can produce entirely different results in different pupils. And a certain experience can produce different results in the same pupil, depending on his or her focus and mood at a given moment.
Incorrect: The participant pays proper attention during fieldwork instructions.
Better: The pupil can make a report on a fieldwork activity which meets the following conditions: …

Be as inclusive as possible when summarising the subject matter in the objective
Do not refer to the content of a particular textbook or lecture transcript. Books and notes usually have the disadvantage that they change over time, with the result that all the objectives have to be checked for correctness every year. In addition the texts of set books and notes can contain ambiguities which hamper assessment. By expressing an objective in the form of a reference to a textbook or set of notes, you are making the objective dependent on the learning method. Users of other textbooks or teaching materials cannot benefit from such an objective.
Incorrect: Study sections 1.2 and 1.3 on page 16.
Better: The participant must be able to list the characteristics of the biotope favoured by badgers.

Use formulations that have an accurate meaning
If an objective is formulated in a way that is subject to interpretation, it will be difficult if not impossible to determine whether the objective has been met.
Incorrect: The participant must write a composition about environmental pollution.

Better: Using examples from the immediate vicinity, the participant must be able to give three examples of pollution and also name three methods of combating pollution.

Strive for simple and clear sentence structure
Objectives can be seen as statements, which can be used to communicate with pupils, fellow teachers, committees and educational researchers about the desired results of educational activities. It is important to ascertain whether the message is clearly understood by the recipient. There is a better chance that someone else will interpret the objective as it was intended, if attention is paid to the clarity of the formulation. If necessary check the clarity of the formulation by asking other people to put the objective into their own words.
Incorrect: Concerning the biotope in a particular place, the participant will be able to name the most important features.
Better: The participant will be able to name the most important features of the biotope in a particular place.

Split complex achievements up into simpler objectives or components
Incorrect: The participant can create and implement a research plan.
Better: The participant can:

- formulate research questions;
- create a research plan;
- collect data for research;
- answer the research questions on the basis of the data collected;
- write a research report.

Use verbs or verbal expressions that refer to observable achievements
An objective only has any point if the achievement of that objective can be determined by the pupil. Therefore it is wise to avoid such terms as:

- knowledge of ...
- know that ...
- have insight into ...
- have a profound understanding of ...
- appreciate ...

These terms can at most be used to express global objectives, for example, in the headers above a series of specific objectives.

Below is an incomplete list of verbs and expressions which do refer to observable behaviour. These and similar terms should be given preference in setting specific objectives.

- name, sum up, draw up a list of ….
- recognise, name, identify …
- distinguish between, indicate the differences between …
- arrange in groups, categorise …
- give a definition of, describe …
- indicate the sequence of …
- explain, give an interpretation of …
- formulate hypotheses about …

- set up a plan for …
- give examples of …
- give an explanation of, name the reasons for …
- make a sketch, plan, overview of …
- describe, report on …
- apply a principle to …

**Specify the conditions under which the desired performance
can be achieved**
The ability of a pupil to deliver a certain performance will depend partly on
the way in which this performance is required from him or her. By
describing several conditions, the performance can be made easier or harder.
Incorrect: The participant can carry out fieldwork activities independently.
More correct: The participant can independently carry out a fieldwork
activity for part of a day based on an assignment and using the materials and
equipment provided.

3.4 Phased plan for specific definition of objectives

Below are the major steps to be taken in the specific definition of objectives.
Each contains a few points for consideration. Following the essential steps
there are some supplementary points for consideration.

Essential steps

It is probably useful to work with a set of index cards. A single objective
definition can be written on each card. Subsequently it is fairly easy to add
or remove cards. The use of index cards also makes it easy to reorder the
objectives in various ways: you simply lay all the cards out on a table and re-
arrange them at will.

You can also use cards to write down survey questions.

a. Making an inventory: what are the objectives of the activity?
Write down a list of all the main objectives of the activity to be evaluated.
Do not make the objectives too detailed at this stage. At a later stage, when
the objectives are more precisely defined and prioritised, there is far less
danger of losing track of them.

It is useful to distinguish between effect objectives and process objectives.

Examples of effect objectives are:

- the participants can name two biotopes which occur in the terrain, and indicate
  which plants and animals will be found there;
- using a map the participants can find their way within the terrain to the various
  locations where they must conduct investigations;
- the participants can make a research plan.

Examples of process objectives are:

- the participants are satisfied with the supervision of the course leader with regard to his/her knowledge of natural science, etc.;
- the participants give a positive assessment of the content of the study week.

A debate can be held within the team about the identified objectives, in which the central question is 'are these the objectives we hope to achieve with this activity?'

b. How clearly are the objectives described?
Look again at each of the objectives that have been formulated. Is it clear?

c. Ordering of the objectives
Prioritise the objectives to find out if you have described all the important ones or if there are repetitions that can be discarded. Points to consider: Are all the objectives equally important? Should all the participants achieve all the objectives to the same extent? Are the objectives of the same order, or are some objectives more general in nature than others are? Is the achievement of some objectives a precondition for achieving others? Is the activity as a whole aimed at all the objectives or are some parts of the activity aimed at achieving specific objectives?

d. What do the objectives relate to: knowledge, skills, values, attitudes, behaviour?
This particular type of ordering can be very meaningful in terms of evaluation, as various types of research instruments are particularly suited to various types of objectives.

Examples:

Knowledge: 'the participants can name the various elements of the water cycle'
Skill: 'the participants can set up a research project, which indicates …'
Experience: 'the participants have a greater reverence for the natural world and feel more involved in it'
Values: 'the participants attribute a higher aesthetic value to the natural environment'
'the participants attribute a higher practical value to the natural environment'
Attitudes: 'the participants have a nature-friendly and environmentally-friendly attitude'.

You may have noticed that the objectives above covering experience, values and attitudes do not comply with the given guideline; they are not formulated in concrete, observable terms. It is very difficult to prepare such objectives without some planning and assistance. You can:

- determine what you yourself think about and what you take notice of, to make you consider how a participant would feel more involved, etc.
- discuss with colleagues what yardstick to use to measure the achievement of these objectives;
- review relevant literature from the target group which might reveal what values children or adults place on the natural world.

e. Over what period of time must the objectives be achieved; do some objectives relate more to the short term, others to the long term?

This method of ordering can have repercussions when deciding when to measure the effects. If certain objectives can only be achieved at a later stage, there is perhaps no point in trying to measure them in the first instance. An instrument, which is used immediately after the completion of an activity, may therefore include different questions to an instrument used in a follow-up survey.

Example:

> It is not unusual in the case of fieldwork for certain groups of participants to make repeat visits. It is evident that in later stages they can work towards more far-reaching objectives, building on what they have already done and achieved. Sometimes, for example, you need to know how the participants use what they have learned in their daily lives.

Further points for consideration

f. Are the objectives in keeping with the wishes and capabilities of the participants?

It may be that the desired effect of an activity will vary from one group of participants to another. It may also be that some objectives are not appropriate to certain groups of participants - for example to a certain age group. Certain groups of participants may also set their own objectives which are too important to be left out of an evaluation, or which should be incorporated into the evaluation as a service to the client.

Example:

> In the case of groups of participants which consist of school classes, the school may formulate specific requirements, perhaps within the framework of certain school subjects, such as biology or geography.

g. To what extent do the objectives fit into the objectives of the organisation as a whole?

If the objectives of the organisation are described, it can be useful to compare these with the objectives of the activity. Do they correspond with each other?

Example:

> If you consider that it is part of the role of the establishment to develop in the participants a feeling of involvement in the natural world, you need to decide whether the activity in question should contribute to this, and if so, an effect objective should be added.

h. How much consensus is there in the team concerning the objectives of the activity?

The team members may not be entirely in agreement about the objectives of the activity. This need not be an obstacle to evaluation. It is, however, sensible to decide how to deal with it. Will the evaluation only cover objectives on which there is consensus, or will all objectives be included, irrespective of the number of colleagues who agree with them?

Example:

> Perhaps the objective of some of the team members is that participants 'learn the most important ecological relationships', 'can set up, carry out and report on a research project themselves', or 'develop a love of nature', while other team members think such an objective is aiming too high. You must then decide whether or not such objectives are included in the evaluation.

i. To what extent do the members of the team actually strive for these objectives in practice?

This can be a difficult question. It can be useful to look separately at the question of the extent to which one really strives to achieve those objectives in practice. This is particularly important for the evaluation phase, since only what has really been attempted can be evaluated.

Example:

> Perhaps all your colleagues agree that participants should learn to design, carry out and report on a research project independently, but in practice they simply cannot devote sufficient attention to it in their working week.
> Tip: at this point you can also consider how feasible certain objectives really are, in other words how realistic they are. After all, some objectives may look OK on paper, but in practice are not attempted, precisely because they are not regarded as realistic.

j. Are the objectives in keeping with national policy?

This is a matter of making a rough check, which at this stage could lead to omitting, adding or reformulating objectives.

Example:

> Educational activities in relation to the natural world and the environment may have to be consistent with the policy of various departments, such as Education, Culture, Agriculture or Environment. The policy of the first department might emphasise certain 'core objectives', while that of others relates to a 'basic ecological curriculum' or 'types of educational objectives'.

k. Are all objectives equally important?

If too many objectives have been formulated ranking them in order of importance is a useful way to sift out a number of less important ones.

Example:

> Teachers think that certain training objectives are fine in themselves but, in the first instance, they set greater store on pupils achieving the 'core objectives', which are largely formulated in terms of knowledge and skills. In this case, you could decide to devote the most attention to the measurement of those objectives.

l. Should all the participants achieve all the objectives to the same extent?

It is possible that the previous question did not lead to a clear distinction between more and less important objectives. If you consider whether all the participants should achieve all the objectives to the same extent, some objectives will immediately be seen to be less important than others.

It may also be that certain objectives are only important for certain groups of participants. On this basis a decision can be taken to include the relevant objectives only in the instrument to be used for that group of participants.

Example:
> Dealing with pupils in secondary education can be complex as participants on various courses come from different schools, from different parts of the country and from different types of education. The initial status of the pupils may vary to such an extent that different objectives are set for different groups.

m. Are the objectives of the same order, or are some objectives more general in nature than others are?
It often happens that some objectives are simply a more detailed specification of other objectives. In that case it is useful to group objectives in that way.

Example:
> A general objective such as 'research skills' can be divided up into specific objectives such as 'being able to set up a research project' and 'being able to carry out an investigation', 'being able to write a research report'. Such 'sub-aims' can also often be further subdivided. In this way one or more 'objective trees' can be built up, which can contribute greatly to giving an overall view of how all the different objectives relate to each other.

n. Is the achievement of some objectives a precondition for achieving others?
The same comments apply here as made in point 8 above. If certain objectives can only be achieved at a later stage, there may be little point in trying to measure them in the first instance.

Example:
> To carry out a research project it may be necessary to be able to use maps. In order to do that it may be necessary to know that there are various types of maps with different uses. To use a map, it is necessary to understand concepts such as scale.

o. Is the activity as a whole aimed at all the objectives, or are some parts of the activity aimed at achieving specific objectives?
Ordering the objectives in this way may show that certain sub-aims are still missing; and they can then be added. If the activity is done in modules, which are used in various combinations, then the objectives of the various combinations can also easily be established. The instrument can also be built up in the form of modules.

Example:
> Often on a fieldwork course different emphasis will be placed on different activities on different days, for example orientation and reconnoitring the terrain, distinguishing biotopes, management work, research.

3.5 Distinguishing objectives for a multiple target group

Suppose that pupils are the ultimate target group, the end users of an educational activity and the teachers of these pupils form a secondary target group.

Aims can then be formulated for both teachers and pupils. In both cases these may be learning objectives such as increased knowledge and skills, training objectives relating to values and attitudes and behavioural objectives relating to changes in behaviour.

The objectives are likely to be different for the teachers and pupils.

A few examples:

	learning and training objectives	behavioural objectives
Teachers	ecological knowledge, teaching skills	better teaching
Pupils	knowledge, skills and values in respect of the natural world and the environment	better behaviour with regard to the natural world and the environment

It is important to make clear distinctions between such different objectives. This is important for the educational activities themselves and for evaluation.

3.6 More detailed specification of effect objectives

Effects in whom?

- in every individual participant;
- in a group of participants;
- the best possible average effect;
- in a specific group of participants.

Effects related to what?

- experience and perception;
- knowledge and understanding;
- skills;
- values and attitudes;
- behaviour.

Effects in comparison with what?

- absolute effects, for example at the end of the course the participants should be able to name the three most significant effects of a particular land use on the wildlife there;
- relative effects, the gain in learning in comparison with a previous time with the same participants;
- relative effects, in comparison with others, for example another group.

Effects when? Effects in the short and/or long term? Only direct effects, or indirect effects as well?

- direct effects are the intended effects and might include, the participants obtain more knowledge or more insight or they develop certain skills, values and attitudes;
- indirect effects might include, new knowledge which becomes manifest in behaviour; for example: you have not only learned why animals must sometimes be left in peace, but you actually do it, both during the course and in similar situations thereafter.

NB: It is a good thing in evaluation to distinguish between criteria and standards:

- criteria largely apply to the characteristics or variables that we are dealing with, about which we collect data;
- standards are the minimum scores which have to achieve before we are satisfied.

Step 4: Deciding when, and from which groups, data should be collected

The selection of the groups

In step 2 we mentioned the following possible objects of evaluation:

• one or more products for certain target groups;
• one or more activities for certain target groups.

The addition of 'for certain target groups' indicates that we are always concerned with specific product-market combinations. Thus for example, an institution may organise study weeks as a type of activity, but if it does so for different target groups, for example, for primary school pupils and secondary school pupils, it will be desirable to distinguish between these when it comes to evaluation. Activities to be evaluated are then: 'study weeks for primary school pupils' and 'study weeks for secondary school pupils'.

In the case of a product we are dealing with users. When evaluating a product one will therefore often try to collect data from those users. In the case of an activity, for example, 'study weeks for secondary school pupils', data will in most cases be collected from the pupils. Data can be collected not only from the ultimate target group of an activity, but also from any supervisors involved. Such supervisors form an extra target group. Finally data can also be collected from those implementing the activity such as the course tutor.

Deciding when to collect data

When evaluating an activity you collect data from the participants after the activity is completed. Even in this simple set-up there is still a choice to be made, namely how long after is 'after'? For example, it may be more important to establish whether the participants still possess certain knowledge, skills and attitudes after some time has elapsed, rather than immediately after the course or activity is finished. If you want to know with greater certainty the extent to which the participants have changed and the extent to which the change can be attributed to the activity, you will have to take several measurements.

With such a slightly expanded regime you can talk about a 'research set-up'. This is a plan which indicates a) when b) about what and c) from whom data are collected.

Section 4.1 below lists a number of options for the selection of a research set-up.

Section 4.2 contains a more detailed description of the methods of reasoning which should be followed, if you wish to draw conclusions about the effects of educational activities. This is mainly about a way of thinking which you

have to master. This does not mean that evaluations must be very complex per se. The most important point is that you must collect sufficient evidence to draw reliable conclusions about the effects.

4.1 Points of measurement, research set-ups and drawing conclusions about effects

How often and at what points should data be collected: once for example, immediately after the activity has finished, or several times?

Often it makes sense to take more than one measurement, for example, taking a measurement before the start of the activity and comparing it with a measurement at the end of the activity. Measurements can also be taken during an activity.

If you want to know whether the participants still possess the knowledge and skills after a passage of time or in other situations, a follow-up measurement will be required. Sometimes you may need to make comparisons between different groups of participants or different locations where the activity takes place.

It is not always necessary to take a lot of measurements. The point is that you have some comparative material.

Effects are often measured by carrying out measurements using the same test before and after an activity. But there are also other options:

- the participants are asked whether, with hindsight, they would have given a different answer in the initial survey if they had known then what they have learned during the activity;
- the people involved are asked in the post-activity survey what they have learned from the activity.

In some cases pre-activity measurements are not necessary, for example when it is very unlikely that the participants could have obtained the knowledge and skills without the activity.

Research set-up with several points of measurement and conclusions about cause and effect.

An effect or result is always the effect or result of something. To be able to speak about the effects of an activity, there must be a causal relationship between the activity and the effect. There is almost nothing so difficult as demonstrating that something is the cause of something else. However, research set-ups have been designed which can serve as the basis for conclusions about cause and effect relationships.

The most traditional set-up is that of a real experiment:

- randomly select a sufficiently large group of people from the population;
- divide this group randomly into two sub-groups of equal size;
- let one take part in an activity;
- measure results in both groups both before and after the activity.

Refinements can be made to overcome some objections to the method, for example having the second subgroup take part in a comparable but empty activity (the placebo), or by missing out the initial measurement in the case of half of each subgroup and carrying out more than two measurements.

In the absence of a control group, it can be difficult to attribute effects to the programme followed, because other factors may also have exerted some influence.

4.2 Aims, effects and conclusions

Introduction

Evaluation of a product or activity is an assessment of what that product or activity is worth. You make the assessment on the basis of certain criteria, the characteristics of the product or the activity being assessed.

To carry out a credible evaluation the assessment criteria must be chosen and publicised beforehand to make sure the correct data is collected and to avoid any temptation to adjust the criteria to suit the results at the end of an activity or project.

Suppliers of educational products and services hope to achieve certain educational objectives with those products and services. And the users of the products or participants in activities also want certain objectives. The quality of educational products and activities is therefore determined by the extent to which they can be used to achieve the desired objectives. In other words the main evaluation criterion is the extent to which the objectives are realised.

Evaluation of an educational activity

We can show the above as follows:

educational activity ➜ realising objectives

The arrow in the diagram means that the realisation of the objectives is the desired effect of the activity. Aims are thus desired effects. The extent to which they are realised only says something about the quality of the activity if the objectives are realised through the activity.

Here we are concerned with demonstrating that the participants have learned something in addition to what they already knew or could already do and that the effect is attributable to the relevant educational activity and not other influences.

Have the participants learned anything new?
Participants in an educational activity are not generally a blank page at the start. They already possess knowledge, skills and attitudes which motivate them to participate in the activity. If you measure only the knowledge of the participants after the activity you do not know how much of this knowledge they already had at the beginning.

The diagram now becomes:

> pre-activity measurement ➜ activity ➜ post-activity measurement

The effect is the difference between the two measurements.

Note 1: It may be that the objective of an educational activity is not to learn something new, but to un-learn something (for example, heavy smoking; fast driving; polluting the environment …) In that case you hope the results of the second measurement are lower than those of the first measurement.

Note 2: Educational activities are aimed at changing people. However, people can easily forget what they have learned and return to their original behaviour pattern. It is sensible to take a measurement not only immediately after the activity but also some time later.

The diagram then becomes:

> pre-activity measurement ➜ activity ➜ post- activity measurement ➜ follow-up survey

Note 3: In switching from the formulation 'the achievement of objectives' to 'the achievement of a difference between the first and second measurements' we have tacitly assumed that the objectives of educational activities are usually relative objectives. They relate to an increase in knowledge, etc., compared with a previous time. We then need to show that the effect can be attributed to the activity.

Note 4: Education is the work of humans and therefore fallible. Thus it often happens that results are disappointing. Does that mean that the activity had no effect? No! The lack of effects of an activity can often be 'explained' by the simple fact that 'the activity was not carried out according to plan'. In other words, we must actually somehow check the extent to which the activity is actually carried out according to plan.

Note 5: Often you want to know not only whether the activity had an effect the one time it was evaluated, but also whether every time the activity is implemented similar effects may be anticipated. This depends on two things:

- the representative nature of the group of participants;
- the evidence that the effect really can be attributed to the activity.

To sum up the above:

- at the very least, you determine the extent to which the objectives have been achieved;
- you make a reasonable case for the activity having an effect;
- you also take note of any other effects whether or not these are expected or desirable.

At a certain point it may become clear what can be expected from the relevant activity in general in terms of effect. Then a *standard* is created, against which the effect of a subsequent activity can be measured.

Step 5: Selecting one or more data-collection methods

When collecting evaluation data you use one or more research instruments. The following types of instrument are often used:

- written questionnaire
- learner report
- knowledge test
- skill test
- logbook
- interview
- observation

Section 5.1 gives a brief outline of these types of instruments. Later in this handbook we provide more detailed directions for the construction of each type of instrument.

So how do you select a suitable method? When should you use a particular method? Are you going to hold interviews, observe, set a test, etc.? There are a number of points to consider when making this choice. These are listed in section 5.2.

Section 5.3 looks at the difference between respondents whom you ask about themselves and informants whom you bring in to make statements about others.

Section 5.4 considers the fact that the very use of a research instrument may in itself have effects, both desirable and undesirable.

5.1 Various types of research instruments

Here we give a brief description of the written questionnaire, the learner report, the knowledge test, the skill test, the logbook, the interview and observation.

We need not go into much detail about **written questionnaires** here; everyone is familiar with them. That is not to say that just anyone can write a good questionnaire. There is quite a range of possible questions, but selections must be made very carefully. For example, the open question, the semi-open question, the apparently open question in which clearly only one word or number can be the correct answer, the question with two response, the question with more than two response options where one or more may be checked, etc.

The **learner report** is a less familiar concept. A learner report asks for the learning experiences of the participants and/or learning effects they have perceived themselves. There are open and closed versions. They can take the

form of open questions (what have you learned?), semi-structured open questions (Complete the following sentences: "I have learned that …", "I have learned how …", "I have learned that it is not true that …", etc.) and questions with closed answers, in which possible learning experiences are formulated and participants can indicate how far they apply, for example: "I have learned why environmental pollution is bad for the natural world".

Open learner reports in which participants formulate learning effects or learning statements, gives rise to various problems:

- delineating what is a single learning experience, for example pupils might write a piece of text in which several experiences are mentioned;
- distinguishing different types of learning experience;
- the validity of the answers may be compromised because people tend to give socially acceptable answers, too many or too few answers. People may not have the language skills to describe their experiences;
- it is too much effort to keep on doing it.

These problems are less evident in the case of closed learner reports. These have repeatedly produced good results.

Like a questionnaire, a **knowledge test** is a familiar concept to everyone. The best known type of knowledge test is the multiple choice. In this case, the incorrect answers must meet a number of specific conditions or the results of the test would not be a good indicator of the knowledge possessed by the respondent. In addition, there are other points to consider like the amount of time the respondents are given to do the test, the difficulty of the test, and the standards which are used to assess what is satisfactory.

Skills are often characterised by effective, observable action and the application of knowledge in practical situations. In a practical test, skills are assessed by means of observation of the action itself. There are also simulations or answering questions about a given case. In addition one can ask people for their own assessment of their skills, or what they have learned in that area.

A **logbook** is a notebook or form which participants or organisers of an activity use to keep track of certain things during an activity or over a specific period. A logbook can be pre-structured with points for consideration, checklists to be filled in, etc. A logbook can be used either to chart the activity being evaluated or to measure certain results.

An **observation** is conscious and goal-oriented. We distinguish between structured and unstructured observation. In fully structured observation it is strictly laid down what must be observed and how. For the measurement of certain skills this could be observing the performance of set assignments. The role of the observer in this type of observation is subject to strict rules. Such observations often provide quantitative information.

Observations can also be useful in obtaining insight into the actual activity being evaluated: what exactly is done, how it proceeds, whether there are problems, how the participants respond, etc. This is often done in an unstructured way. Unstructured observation means carrying out more qualitative research from a relatively wide range of perspectives.

5.2 Checklist of criteria for the selection of a research instrument

Below is an elaboration of the criteria used in selecting a research instrument. Each type has its own advantages and disadvantages.

Usability

The method must be useable. For example, in interviews you are very dependent on the skills of the person doing the interviewing. With regard to suitability for the target group, remember that a written questionnaire cannot easily be used with young children.

Data collection and processing time

This is about the amount of time you devote to the collection and processing of data. To obtain maximum information from the greatest number of people in the shortest time, by far the best method is the written questionnaire. With regard to the data collection time, you need only add an extra address and make an extra copy of the form for each extra respondent.

With an interview, both the collection and the processing of the data takes more time, whereas with logbooks and learner reports it is mainly the processing which takes time.

Reliability of data

Reliability refers to the extent to which a repeated measurement in respect of the same respondent produces the same result. For example, if there are two observers, it is quite likely that the results will vary widely. Observers see things from their own frame of reference and are therefore selective and subjective. Reliability is also important in questionnaires. If the questions are not clear, people will interpret them differently.

Validity of the objectives to be evaluated

The selection of a research instrument is determined by the objective to be evaluated. The development of values cannot be measured using a test and an increase in knowledge cannot be measured by observation. If you are dealing with objectives relating to behaviour, observation is usually more valid than a questionnaire. What you say and what you do are two different things.

However, sometimes you are dealing with forms of behaviour which do not often arise. Therefore you have to ask about it. So the question arises as to whether you are measuring what you want to measure. One way to determine validity is to use different instruments in conjunction with each other and then compare the data collected.

Validity is particularly important in instruments which deal with self-reporting. Particular question marks can be placed against answers to questions about attitudes, values and behaviour, as people naturally tend to give socially acceptable answers.

Contact with the respondent

Holding interviews has certain advantages. The interviewer can continually adjust the course of the interview, avoid incorrect interpretations, clarify the situation, ask follow-up questions, etc. However, a disadvantage can be that the contact between interviewer and respondent can affect the responses. Training in interviewing techniques will help ensure that the personality and interviewing method of the interviewer do not adversely affect the responses.

A questionnaire does not have the advantages of an interview, but it does have the disadvantages, albeit to a lesser extent. The interviewer is not present in person, but is represented by the questionnaire. The tone of the questions and the layout and presentation can influence the answers.

You will see from the above that many factors play a role in the selection of an instrument. It is possible that not all the objectives of an activity can be measured with a single type of instrument. In that case you can use several methods.

5.3 Working with respondents and informants

A respondent in an interview is someone about whom we wish to know more. Respondents report about themselves and their own situation. An informant is someone who remains outside the scope of the research, but is asked to cooperate in reporting on others or on situations in which the informant is not personally involved.

A similar distinction is that between subjectivity and objectivity. In many cases you can ask for a 'subjective assessment' or for 'objective information'. For example:

- **Subjective:** sometimes - from time to time - reasonably often - often
- **Objective:** never - roughly once per year - roughly once per month - roughly once per week - practically every day

The first formulation asks about *the perception* of the frequency by the respondent. The second formulation asks more objectively about the *actual* frequency. A researcher could also start by asking for the objective information and thereafter the subjective ('*do you think that's often, or not?*'). This could be done by using a matrix of two questions side by side.

5.4 Desirable and undesirable effects of the use of a research instrument

Measuring someone's knowledge or attitude can affect that knowledge or attitude, and hence the measurements themselves. For example, if a questionnaire contains questions about the values which the respondent considers important, this may cause the respondent to think about it, and then become more aware of certain values and give different answers to those he or she would initially have given spontaneously.

In principle you should be able to measure a concept or variable in a number of ways, without affecting the research results. For example, if the results of a written questionnaire show clear differences from those obtained by interview or observations, one must normally conclude that there is obviously some mistake in the data collection. The measurement results are then too heavily dependent on the type of instrument used, and the instrument therefore exerted too great an influence.

The research instrument, or the activity of collecting data can also directly affect the result of the measurement:

- it may be that the situation is highly unusual for the respondents so that they are not at ease or they do not have the skills to provide proper answers to the questions;
- another example is that people often think that in a survey of attitudes, there are right and wrong answers, and so perceive it as a test situation;
- yet another example is that people are set off thinking by the data collection process and partly as a result, change their views;
- a final example is the possibility that questions or assignments are formulated in a way which not everyone understands or interprets in the same way. Attention must therefore be paid to the frame of reference for the respondents, in other words their knowledge, culture, etc.

In general instruments can often be used to stimulate discussion for example in a school team to stimulate change etc. Experience shows that, for example, filling in a questionnaire together can help to set useful developments in process.

Step 6: Listing one or more factors which facilitate or inhibit the achievement of the objectives

With an educational product or activity you are trying to achieve certain objectives. To achieve those objectives the product or activity will have to have certain characteristics and you need to decide what these are. If the characteristics, factors and conditions which are important for obtaining the desired effects can be influenced by the educational establishment, this knowledge can be used in the preparation and design of the activity.

There are four groups of factors which matter:

- characteristics of the product or activity as designed;
- the way in which the product is used, or the activity is carried out;
- characteristics of the users of the product, or participants in the activity;
- the circumstances under which the product is used, or the activity is carried out.

Step 7: The creation of one or more research instruments

In Step 5 a research instrument was selected. Whatever selection you make, it will always be necessary to draw up a list of questions, items or checklists. We have already looked at the content of these in Step 3 - aims and in Step 6 - influential factors. Our concern now is to create the instrument.

Section 7.1 deals with different levels of measurement.
Section 7.2 describes quality criteria including reliability, validity, acceptability and usability.
Section 7.3 explains working with scales.
Section 7.4 describes ways of developing and testing a research instrument.

In addition each type of instrument has its own advantages and pitfalls. The quality of an instrument depends on the extent to which these are taken into account. For more on this subject, see the more extensive treatment of the various types of instruments in the Appendix I-3 in this handbook.

7.1 Different levels of measurement

The level of measurement influences the information a question elicits, and the way in which the response can be analysed. We distinguish between the following four levels of measurement.

Nominal

This refers to placing the answer in one of a number of categories which differ from each other in terms of quality. Familiar examples are questions relating to church or political party affiliation. You either belong in a particular category or you don't in terms of measurement.

An example:

What method do you use for teaching biology?

      0   method A
      0   method B
      0   method C
      0   etc.

Ordinal

The question relates to a concept or characteristic where there is more or less. An example is the degree of satisfaction; one person is more satisfied than another, or a person is more satisfied with one course than another.

An example:
Please indicate below how often the course tutor carried out the following
activities:

	never	sometimes	regularly	often	always
gave an explanation	0	0	0	0	0
gave examples	0	0	0	0	0
had the participants	0	0	0	0	0
practice	0	0	0	0	0
etc.	0	0	0	0	0

Interval

Here the distances between points on the scale are equally great. A familiar
example is the measurement of temperature in degrees Celsius. The
difference in heat between 5 and 10 degrees is equal in size to that between
10 and 15 degrees.

An example
How often have you applied what you have learned in the last month?
   0 never
   0 once
   0 twice
   0 three times
   0 etc.

Ratio

This relates to questions where there is an absolute zero. An example where
there is measuring at ratio level is the measurement of velocity. It is
meaningful to refer to a velocity of 0, while a velocity of 10 is twice as fast
as a velocity of 5.

It is important to distinguish between these levels of measurement, since it
will influence what you can and cannot do when you are processing the data.

The boundaries between the types can be blurred. For example, ordinal data
are often treated as if they were interval data. This is usually justified as long
as the 'distances' between answer categories can be considered equal. For
example, it is not true of 'never - sometimes - often' but it is true of 'never -
sometimes - regularly - often'. Another case is the variable with only two
values, for example yes/no or male/female. Such variables are sometimes
called **dichotomous** variables. Such variables can be interpreted as a very
simple form of an ordinal or interval measurement level. In other cases a
dichotomous variable can best be seen as a nominal variable with only two
categories.

Some examples:

- a question about a fact - yes/no;
- a statement or proposition - agree/disagree;
- what is your gender? - male/female;
- are you married or unmarried? - married/unmarried;

The level of measurement of a variable is not an absolute fact. For example, one can place the values of a quantitative variable in a limited number of classes and, as a result the variable, can be better accommodated in a table.

An example:

What is your age?
- 20 or less
- 21 - 25
- 26 - 30
- 31 - 35
- 36 - 40
- 41-45
- etc.

7.2 Quality criteria for research instruments

Two important quality criteria which must be met by research instruments are reliability and validity.

The **reliability** requirement means that you must expect that if the measurement is repeated, similar results would be obtained. This can be investigated in a variety of ways. The methods all rest on the same principle. If I have several instruments of measurement, for example several questions in the same questionnaire which are thought to measure the same thing, then the responses should be similar or closely interrelated. This principle is also used in the construction of scales.

A research instrument is **valid** if it measures what it is thought to measure. For example, if you wish to measure a certain theoretical concept such as self confidence then the questions must measure that concept.

There are also a few other important requirements for the practical application of research instruments.

Under **acceptability** we can include requirements such as ensuring that the research does not manipulate respondents, discriminate against certain groups or violate personal privacy.

A specific aspect of **usability** is **discrimination**. For example, if you want to investigate the effects of a course on the knowledge of the participants, you will glean little information from questions which are far too easy or far too difficult.

7.3 Working with scales

In research, data are often collected on a relatively large number of variables. It is not easy to report comprehensively on them, particularly if the interrelationships must be described. For this reason the available data are usually first condensed into a smaller number of variables. This is taken into account in the construction of the research instrument.

There are also other reasons for condensing data into a small number of variables.

First, the research can cover many aspects to enable respondents to give the full picture, rather than just part of it. Each aspect should be categorised in a different item.

Secondly, experience shows that responses to individual items are unreliable because people read carelessly or make a mistake when ticking a response option or are guided by associations which evoke a certain formulation. This unreliability occurs when people are asked to give the same information again. This problem can be solved by combining the responses to several items.

Thirdly, by combining a number of items, the number of possible scores increases. A score for one item with 6 response options can range between 1 and 6. The more items the more accurately you can measure, or the better you can discriminate between the respondents.

The following is an example of a desirable scale. It relates to a series of items included in a written questionnaire for head teachers and teachers in primary education, and is intended to measure personal involvement in the natural world and environment. The possible responses were 'disagree' (coded and entered as 1), 'agree slightly' (2), 'agree to a certain extent' (3), 'agree to a large extent' (4). The percentages shown in the table are the totals of the percentages of the responses 'agree to a certain extent' and 'agree to a large extent'.

A different form of scale analysis must be used to determine whether a series of items measures the same concept. A much-used form of scale analysis is item analysis. Briefly, the background to an item analysis is as follows.

If the responses to a number of items are fairly strongly interrelated, those items probably measure different aspects of the same concept. The interrelationship between the items is calculated by means of correlations. A correlation can vary between +1 (a perfect positive relationship; an example: the height of people in centimetres and inches) and -1 (a perfect negative relationship; an example: the height of people and the distance between their heads and the ceiling (of the room in which the measurements are taken). Significant correlation is normally taken as lying between +/- 0.35.

	Head Teacher agrees to a certain extent	Teachers agree to a large extent
I read natural science and/or environmental magazines and/or newsletters published by nature and environmental organisations	66%	61%
I often read newspaper articles concerning nature and the environment	80%	9%
I regularly watch/listen to nature and environment programmes on TV and radio	62%	70%
I discuss with others how they can contribute to a cleaner environment	53%	52%
I try to convince other people to behave in a way which is better for nature and the environment	57%	57%
I comment if other people behave in a way which is environmentally-unfriendly	54%	53%
I know a lot about the world of animals and/or plants	49%	53%
I love the countryside, I like walking in the country	87%	90%
I regularly go out to watch wildlife	26%	28%
On walks, I collect stones, feathers, shells, seeds or fruits	22%	31%
I like to identify plants, fossils, etc.	18%	18%

7.4 Developing and testing a research instrument

When developing a questionnaire make a draft first and then try to improve on it by:

- having experts assess the draft;
- having representatives of the intended respondents assess the draft;
- trying out the draft on a limited sample of intended respondents.

To test an interview you can run pilot interviews.

There are several reasons why adjustments to questions might be needed:

- they cannot be understood by the target audience;
- they are ambiguous and people interpret the questions differently;
- there is a disappointing response to the questions as a whole. This can indicate that the questionnaire is too long, the subject is not interesting, the purpose unclear, the wrong time has been chosen, or any number of other factors that might need to be investigated.

Step 8: Deciding who to include in the investigation

Step 4 required a decision about the group or groups from whom data would be collected. However the choices are made, you will often decide not to collect data from all the people involved, but only from a **sample**.

There are two important considerations here:

- how do the advantages of a small research group stand up against the disadvantages of taking and dealing with a sample?
- how will the sample be taken, and how large must the sample be, before you can obtain data which is sufficiently accurate and sufficiently representative?

Section 8.1 deals with working with samples in more detail.

8.1 Working with samples

Introduction

Taking a sample rather than the whole population or target group can make the research cheaper and faster to implement without prejudicing the quality. When working with samples, the absolute number of respondents is more important than the percentage. For example a random sample of 10 teachers from a group of 30 can give a good picture, but you do not necessarily need 30 teachers to obtain an equally reliable picture of a group of 90 teachers. We give this example to illustrate that the absolute number is more important. In the type of research we are dealing with here, it is not generally necessary to work out precisely how large a sample must be to achieve results of a certain degree of reliability or accuracy. In practice, a sample is rarely fully random, and because there is often a degree of non-response, there is not usually much point in making such calculations.

It is also possible to combine sampling and non-sampling. For example, you can ask everyone to fill in a questionnaire and then use interviews with some of the respondents to go into certain matters more deeply.

Sampling can apply to time and space as much as people. For example, a sample of three weeks could be taken from all the available weeks.

Types of sample

Three methods of obtaining a sample are looked at below:

- *a simple random sample:* this is an indiscriminate sample and in practice truly random samples are rarely ideal;

- *a stratified sample:* where a population consists of different groups or 'strata', for example, small and large schools, or students in different year-groups;
- *a multi-stage sample:* the following example illustrates the concept. Schools might first be sampled, then the teachers within the school.

By taking a random sample from each group in a stratified sample it is possible to obtain sufficiently accurate results from a relatively small overall sample. There are various ways to obtain a stratified sample. One way is to take an equally great *percentage* from each stratum. This is also referred to as 'proportional' sampling. An advantage of this method is that the results of all the strata can be combined and conclusions can then be drawn about the entire population without the need for further processing. Another possibility is to take an equal *number* from each stratum. In this case, before any conclusions can be drawn about the population as a whole, it is necessary to weight the results in each stratum to prevent bias.

Different types of errors and bias

Various errors can happen when using questionnaires with sample groups:

- *sampling errors*: these are distortions which can occur in samples because the sample is not representative of the population as a whole;
- *representation errors:* these occur when the framework of the sample such as the list or file with names and addresses, does not coincide with the target population;
- *distribution errors:* these are administrative errors in the sample framework, such as incorrect addresses, duplicate entries, etc.;
- *non-response:* many people do not return questionnaires, so the sample ceases to be representative;
- *measuring errors:* these result from flaws in the questionnaire.

Note:
The following three concepts often cause confusion:

- *randomness*: this is a feature of the *procedure* of taking samples;
- *representative*: this is a feature of the *respondent group*, compared with the overall population;
- *generalisation*: this is a feature of the conclusions; they may be applied or generalised to the whole population.

Step 9: Collecting the data

We can distinguish four groups of research instruments:

- written instruments, which are handed personally to the people who are to fill them in;
- written instruments, which are sent to the people who are to fill them in;
- interviews and observations;
- working with assessors.

Written instruments by post
In this case no complicated organisation is required. There must be sufficient copies available, brief instructions, space, time and quiet to fill them in. Respondents may be asked to put on their names.

Written instruments by post
When people are contacted by post many will not respond. It is necessary to judge the extent to which the data obtained are sufficiently representative. There are various ways of doing this, some more laborious than others.
In section 9.1 there is a detailed review of the organisation and logistics of written instruments. Section 9.2 describes various ways of dealing with non-response.

Interviews and observations
People have to be employed to carry out interviews and observations. Points for consideration here are:

- the number required naturally depends on the number of interviews or observations to be held and the way in which they are organised. It is preferable not to involve too many people. The more interviewers or observers are used the more difficult it becomes to ensure that they work in a comparable manner;
- giving the interviewers or observers instructions and, where necessary, training;
- making and confirming in writing appointments with those to be interviewed or observed. When confirming, an explanation is given concerning the purpose of the interview or observation;
- dealing with problems, for example, if an interviewer falls sick;
- checking the delivered interview or observation protocols. Are they complete? Is there a name on them?

Working with assessors
When working with assessors, similar points arise as mentioned above.

Section 9.3 deals with the specific options for data collection in a situation where there is a multiple target group.

9.1 Organisation and logistics for the use of written instruments

Approaching respondents

Written questionnaires should be accompanied by a letter in which you introduce and explain the purpose of the research. A letter is often also sent when using interviews, in which the appointment is confirmed, further information is provided about the research and the subjects to be discussed are listed so that the respondent can prepare for the interview. The letter can indicate that the respondent may expect to receive a phone call in the near future asking for their cooperation.

A prior announcement of a research project can also sometimes be placed with good effect in a medium regularly consulted by the target group, such as a specialist journal, a notice board, etc.

A letter accompanying a written questionnaire might include:

- the name of the organisation carrying out the research;
- the name of the client if applicable;
- the subject of the research;
- the reasons for the survey;
- the objective and significance of the survey;
- the reasons for approaching the recipient at this time;
- what is attached, such as a questionnaire and reply envelope;
- the request to complete the form;
- what the recipient is asked to do. The return deadline is usually 3-4 weeks;
- the use which will be made of the data;
- the guarantee of anonymity and explanation if a respondent number is used;
- the undertaking that, by way of thanks, a summary of the results will be sent to the recipient if desired;
- the opportunity to ask questions or obtain further information.

A covering letter must be developed just like the questionnaire. Ask other people to read the draft and comment on it. Improve it until the text is sufficiently clear and attractive.

It is advisable to be as open and honest as possible about the objective and set-up of the research to guarantee anonymity and to promise to use the results prudently. Honesty is the best policy and often also produces the most valid information. It is also advisable to put the instructions for filling in the questionnaire, not in the covering letter, but on the questionnaire itself.

It is equally important when approaching participants using other forms of research instruments to pay attention to the points made about the covering letter for a questionnaire. It is best not to offer any incentive other than reporting back on the findings. People can react very negatively to incentives.

Some logistical aspects and reminders

- establish what is to be sent out;
- agree a return address;
- select the sample;
- brief the person dealing with questions once the questionnaire is sent out;
- log the returns;
- send out reminders, a reminder can generate a further 20-30% response.

Possible extras and variations
The process will be modified if you are using intermediaries. For example, a teacher might get the pupils to fill in a questionnaire for you. You may send a pack of questionnaires, or come to hand them out personally.

9.2 Dealing with non-response

A problem with research instruments like written questionnaires is non-response. That is, not all the questionnaires come back. This makes the data biased in favour of respondents and therefore less reliable. A response of 70% or more is good but if the response is less than 60%, the researcher should attempt to discover the extent to which the response is representative of the whole sample.

Ways of dealing with non-response

To find out how representative the responding group is, you can carry out a 'non-response investigation'. You choose certain questions from the questionnaire which are crucial to the research and to which the respondents have given varying answers. Next you take a sample (say, 20%) of the non-respondents. These people are phoned. It is advisable not to mention that they haven't returned their forms. Restrict yourself to asking the selected questions. Next you compare the distribution of the answers to the questions in the non-response investigation with the distribution in the group which did complete and return the forms. If there is little difference, you know that the response group is representative of all the participants. If there is a difference, you can correct the results of the questionnaire accordingly. For example you can interpret them differently, or you can raise or lower them by a certain percentage.

If the non-respondents are clearly different from the respondents, then you could argue that respondents who send in replies later are more like people who do not react at all than people who reply immediately. By keeping a record of the date replies are received you can determine the relationship between the answers to the main questions and the speed of response. If they are unrelated, you can argue that the response is representative. If they are related, you can correct accordingly. This test could be applied to those who only reply after a reminder.

Partial non-response and serviceable response

Not all respondents answer all the questions so for each question there is an extra element of non-response. The comments above about being representative are thus also important for each question. If certain respondents have left a lot of questions unanswered or illegible, it is better to remove them from the database altogether.

9.3 Ways of collecting data from a multiple target group

There are extra options for collecting data from a multiple target group:

- teachers can be brought in to collect data from pupils;
- teachers can be asked to assess the activities and learning results among the pupils;
- pupils can be asked to assess the activities of the teachers;
- instruments for measuring learning results among pupils can be given to teachers so they can evaluate the learning results of their pupils themselves.

Step 10: Processing and analysing the data

A **plan** of **analysis** must be made on the basis of the research question and the type of data collected that is a brief summary of what you want to know and what procedures are required to obtain the necessary information.

There are six possible types of procedure:

    a. Summarise qualitative data in a text. All the responses to open questions are read through and then summarised as well as possible. This may sound easy, but it is anything but.

    b. Categorise open question responses. The responses to open questions are converted into a code which stands for the categories into which the responses fall. To do this you must first devise a good set of categories.

    c. In further analysis of quantitative data the first step is always to count the results for each question in each answer category. This 'tally' can be kept manually, usually as percentages, although sometimes means are also calculated. This type of descriptive analysis can be carried out without using a computer. However, a computer will greatly accelerate the process.

    d. The next step is to determine the extent to which certain groups together form a reliable scale (see also 'Working with scales' in step 7). This is best carried out by computer.

    e. In the case of certain data, you can investigate the differences between two or more groups of respondents, for example participants in week A and participants in week B, girls and boys or measurements. Here you can make do with a descriptive account of the differences, for example using the percentages or means for the various groups. You can also carry out statistical tests to determine whether the differences are significant.

    f. You can also investigate the extent to which the data relate to each other. For example, if the age of participants was a factor in the results. Here, too, you can make do with a descriptive account of the relationships or conduct statistical tests to determine whether the relationships are significant.

10.1 Categorising and recording answers to open questions

In research you often work with open questions. Such data are often processed qualitatively in the form of a descriptive summary of the answers.

Tip: this can be facilitated by putting the answers for each respondent in the computer and providing them with codes for the question and the number of the respondent. In this way all the answers to the same question end up next to each other.

It can be unsatisfactory to make do with a qualitative, narrative treatment. The fact is, this is a weak sort of research, from which relatively few conclusions can be drawn. The main methodological problem of a purely

narrative treatment is verification. There is a high level of subjectivity. It is very easy to glean from a number of statements what strikes you or what you think is important. But something more is required of good research. Here are a few options.

Use the available data to see what content categories can be distinguished within it and then allocate codes to the categories. Next the answers of the respondents to the various questions are recorded in each category and possibly converted into percentages. Once you have created categories, you can enter the data in a computer for further processing. A great advantage of this is that you not only have access to the number of times the various answers occur, but you can also easily relate various data variables to each other.

An example:
> In the autumn of 1994 a 'stock-taking' survey was conducted into extra-curricular educational activities relating to the natural world and environment. Of the 500 or so organisations which were approached, 150 returned completed questionnaires. Various open questions were included in the questionnaire. One of these was a request to describe the most characteristic activity of the organisation. The many written answers were divided into a number of categories:
>
> 1. developing and producing materials;
> 2. providing information and facilitating communication;
> 3. developing and running courses, training sessions or lessons;
> 4. providing supervision and advice;
> 5. organising outside activities;
> 6. other.

A few comments/pointers

- the number of categories should be neither too small nor too big: 5 - 8 is usually optimal;
- an 'other' category is unavoidable. This category should not get too big otherwise the main categories no longer convey much meaning. At least 80% of the answers must fit into one of the main categories;
- the main categories must, as far as possible, carry equal weight;
- the numbers 1-6 are the values of a nominal variable These values can be used for further qualitative processing.

In the same survey was the following open question: 'What do you currently consider to be the greatest barrier to implementing extra-curricular nature and environmental education or related activities?'

Many respondents mentioned more than one barrier. Most answers could be put into the following categories:

- lack of time and money;
- lack of and high turnover of volunteers;
- lack of interest among target groups and membership;
- general decrease in interest in the environment;
- lack of continuity, or too much inconsistency on the part of the government;
- lack of clear structure in the field, or unregulated growth;
- the broadness of the field and the current lack of professionalism.

Because many respondents mentioned more than one obstacle, the answers could not be accommodated in a single nominal variable. Each of the above categories therefore became a dichotomous variable in itself. Each respondent was given a score for each of these variables, namely: 0 (most common) for 'not applicable' and 1 for 'applicable'. These scores can also be entered for further processing if desired.

Recording and taking percentages

Even without entering the data in a computer, categorised data can be quantitatively processed, simply by tallying. It does not mean much in itself to say that, for example, 69 respondents gave a particular answer. Whether that is a lot depends on the total number of respondents who took part in the survey. For this reason percentages are usually calculated.

The results of recording and taking percentages might look like this:

main type of activity	organisations of type X		organisations of type Y		total	
	n	%	n	%	n	%
1. development and production of materials	25	28	10	17	35	23
2. provision of information and/or facilitating communication	15	17	20	23	35	23
3. development and/or running of courses, training sessions or lessons	20	22	5	8	25	17
4. provision of supervision and/or advice	10	11	10	17	20	13
5. organisation of outside activities (guided tours, excursions, etc.)	5	6	0	0	5	3
6. other	15	17	15	25	30	20
Total	90	101%	60	100%	150	99%

The letter 'n' stands for the total number of respondents.
Due to rounding up, the percentage total is not precisely 100%

We see that organisations of type Y show a different pattern from organisations of type X. In particular we see more 'information and communications' and more 'other' and less 'courses, training sessions or lessons'.

10.2 Making a plan of analysis

Data should not be analysed in a hit and miss way but rather on the basis of what you want to know. For example analysis can be aimed at investigating how often certain answers are given; usually this is expressed in percentages. Averages and the degree of distribution of answers can also be determined. You may also want to investigate the differences between certain groups, or the degree of correlation between certain data. Ultimately the researchers may want to be able draw conclusions themselves about the degree of change in a target group or the impact of a certain measure such as an educational activity.

Possible questions are:

- *descriptive results:* How many? How often? How many percent? What mean? What distribution?
- *scale analyses:* Do these items together form a reliable scale? What aspects belong to this concept? Can this concept be reliably measured? How reliably has this concept been measured?
- *analyses of differences and relationships:* Are there differences and if so, how great are they? Are there relationships?
- *interpretations of results, conclusions:* Have changes taken place? Have there been any effects?

Next we will look more closely at different types of analyses.

Examples of descriptive analyses:

- the number of respondents who give an answer;
- the percentage of respondents who give an answer;
- the mode or most frequently chosen value;
- the mean or average;
- with such variables, it is also common to determine how the answers are distributed about the mean; this is usually done by calculating the standard deviation;
- a graphic representation can also give a good impression of the distribution of answers.

Scale analyses

If a scale is sufficiently reliable the answers to individual items can be combined to arrive at a single score, usually by calculating the mean. The mean then represents the entire concept in question.

Analyses of differences and relationships

The third group of research questions we look at here concerns differences and relationships. For example, if you want to find out the difference in managerial skills between men and women, you look at the relationship between the variable 'managerial skills' and the variable 'gender'. An

analysis of differences and relationships can be described and tested statistically. In the latter case the question is whether the difference or relationship is so clear that it cannot be put down to coincidence.

Questions relating to the interpretation of results

Conclusions about changes and effects can only be based partly on analyses. It is more a matter of interpreting the results, for example, concerning a difference between two measurements:

- has anything really changed?
- can that change be attributed to the activity?
- how can the changes and effects be explained?

Step 11: Interpreting the results and drawing conclusions

At the end of a research project or evaluation the last question is 'what do I know now?'

To what extent are the data themselves sufficiently reliable and valid?
The value of the data obtained stands or falls on the quality of the research instrument used. In practice there is often a weak link here, and people are far too inclined to draw far-ranging conclusions from answers to what were in fact 'poor' questions.

In statistical terms, how significant are the differences and relationships found?

Three aspects play a role in assessing a difference or a relationship:

- how great is it? For example, how much do two means differ or how great is the correlation?
- to what extent does it actually represent something in statistical terms or could the difference or relationship be a matter of coincidence? In other words, if you repeated the experiment you might not find any such relationship;
- how significant is it, does it convey any meaning in practice?

It is important to determine the extent to which differences and relationships are statistically significant, so that you do not jump to the wrong conclusions.

To what extent do the results indicate actual changes or effects?
The question of whether there are actual changes can only be answered on the basis of an assessment of the quality of the instruments used. Only if they are sufficiently reliable and valid can differences in score between two measurements be taken seriously. If, for example, certain questions in a questionnaire can be interpreted in different ways, some participants will understand them one way, some another.

The question of whether the desired effects have been achieved goes a step further. Here it is a matter of whether the changes which can be established can also be attributed to the educational activity in question. Such a conclusion can only be drawn if other possible explanations can be excluded.

What conclusions can now be drawn, and what recommendations made?
You must not get bogged down in all the data, but ultimately draw conclusions - which is in fact the entire purpose of the exercise.

11.1 Establishing effects

There are other ways of measuring effects apart from the pre-activity and post-activity surveys described above:

- the participants themselves are asked in the second survey if, with hindsight, they would have given a different answer to a question before the activity. This is particularly useful, for example, when measuring skills in which there is a good chance the participant will rate their skill higher before the activity than it proves to be during the activity. For example: 'I always thought I was good at finding my way around in unfamiliar surroundings, but when I had to find my way in the woods with a map and a compass, I really wasn't very good at all.'
- in the second measurement the people involved are themselves asked what they have learned;
- other people such as the project leaders and supervisors are asked to estimate what the participants have learned.

Step 12: Writing a report on the evaluation and the results and conclusions

A research report should in principle be complete. However, that is not to say that you have to describe everything. Not only would that entail a lot of work, but the report would be unreadable. The art is to summarise the design, implementation and results in such a way that no essential information is left out.

The research must be verifiable. The reader must be able to discover how you reached your conclusions. Information about your own working methods which is not relevant, is omitted.

In a research report you should always make a clear distinction between the results themselves and the interpretations, consequences and recommendations which you as a researcher attach to them.

A report will have different forms for different purposes and target groups.

Target Group	Form	Purpose
client	- report - oral presentation (with acetates and other media)	- fulfil the contract - promote the use of the results
respondents/informants	- written summary of results (possibly with comparison of overall view and tailored to recipient)	- returning the favour - as a form of feedback
one's own organisation, colleagues, colleagues elsewhere	- written summary - presentation and discussion - article in technical journal	- exchange of information - consultation - signalling developments - presenting possible applications
scientific forum	- article in a scientific journal - paper at scientific conference - electronic publication	- contribute to science - working on own list of publications - conduct discussion - making contacts
general public	- press release - article in paper	- bringing to the attention of others - spark off debate

12.1 Suggestions for the compilation of an evaluation report

The following are important points to consider:

- length?
- what information will you choose to report?
- will you use the same report for everyone?
- what style will you use?
- how will you disseminate the findings?

Various objectives must be combined in a research report:

- the results of the research;
- the design and implementation of the research must be described and justified;
- the report should include the research questions and answers. In other words, at the end of the report you always return to the beginning. What was the purpose behind the research, and what do we know now?

In a research report a clear distinction must also be made between the results in themselves and the interpretations which you, as a researcher, attach to them. For this reason a report often has a separate chapter for 'discussion and recommendations'.

Over the years a number of conventions have developed for the layout of research reports. Such a standard layout may make the report a little dry, but the advantage for researchers is that they cannot easily omit something, and for the readers that it is easier to find the information they want.

Below is a summary of the sections which may make up a report:

- title page
- table of contents
- introduction
- summary
- introduction / bookmark
- formulation of the problem / background information
- literature (theory; available data from other research)
- design and implementation of the research
- results
- conclusions
- discussion and any recommendations
- bibliography
- any appendices

Where do you draw the line between results and conclusions? There is a gradual transition. You could make the following distinction:

- small conclusions, which you link to separate tables or results of analyses and which are formulated in pieces of text in a chapter called 'Results' . For example, 'We can conclude from table 5 that …';

- big conclusions, which are more all-embracing or summarising in nature, in which an answer is given to the research questions and which are placed in a chapter called 'Conclusions'.

Many readers want to use the table of contents of a report to see roughly what the report is about and where to find what. Therefore do not make the table of contents too brief. Let it give an overview so that the structure of the chapters and sections can be seen at a glance.

Important: Always check that the titles of chapters and sections in the table of contents correspond to those in the text, and likewise that the page numbering in the table of contents corresponds to the page numbers in the text.

Chapters, sections and paragraphs serve an important function in the arrangement of a report. A good lay-out can help you to work systematically during the writing process. Do take care that a section consists of a rounded whole and that a heading also covers the content. It is very confusing for the reader if a section on taking samples suddenly comes up with information about the construction of a questionnaire.

There are no rules about how a chapter should be broken down into sections. While it is a good habit to divide a chapter of a certain length into sections, the reader may have difficulty keeping track if there are a great deal of sections and subsections.

Paragraphs also have an important function. Good use of paragraphs makes it easier to follow the argument. Do not introduce a new subject at the end of a paragraph. Ensure that a paragraph is a rounded entity in terms of its subject.

Step 13: Reaping the benefits of the evaluation

To reap the benefits of the evaluation:

- ensure that the research work is well rooted in the organisation;
- believe in your research work and continually stress the importance of quality requirements;
- do not overestimate the reliability of your data or the ability to generalise from it;
- deliver the report on time, neatly and in accordance with any agreements made;
- give an explanation.

Creating and presenting a report can promote and support debate about the use of the research results.

At the very least the results of an evaluation should be discussed by the people involved and interested parties.

It is a good idea to look at the possible consequences of the results, and to do this jointly. Research or evaluation results are the perfect material for discussion in a team or organisation.

Research has several possible consequences. Depending on the results these could include:

- continuing on the current course or with the current plan;
- implementing changes;
- ceasing a certain activity;
- developing something new;
- the learning effect on yourself;
- encouraging target groups to make better use of the services on offer;
- justifying yourself to the funder, the client or the community.

Generally speaking, research results can be used in a variety of ways:

- *instrumental use:* certain decisions are based directly on the results;
- *conceptual use:* the results contribute to advancing understanding and influence people's thinking. This is the sort of use that is generally the best to aim for;
- *political use:* the research results are used to reinforce certain positions or influence developments;

To maximise the effect, it is often desirable to present the report and explain it in person.

A report can also be distributed among third parties, because it is a requirement, or as a showpiece, or simply for their perusal. However, bear in mind that the impact of distributing written material is often disappointing. In many organisations incoming information has difficulty getting through the system, while the intended recipients do not always read it, let alone act on it.

Appendices

Step 1
Determining the reason for the evaluation, or the purpose(s) which the evaluation is to serve

Step 2
Determining the object of evaluation

Step 6
Listing one or more factors which facilitate or inhibit the achievement of the objectives

Step 3
Defining the objectives of the product or activity

Step 4
Deciding when, and from which groups, data should be collected

Step 5
Selecting one or more data collection methods

Step 7
The creation of one or more research instruments

Step 8
Deciding who to include in the investigation

Step 9
Collecting the data

Step 10
Processing and analysing the data

Step 11
Interpreting the results and drawing conclusions

Step 12
Writing a report on the evaluation and the results and conclusions

Step 13
Reaping the benefits of the evaluation

Appendix I: Guidelines for creating research instruments - Step 7

Certain general factors play an important role in the creation of an instrument. First of all the parts of the instrument, for example the questions in a questionnaire, differ in terms of **levels of measurement**. Sometimes there are only two possible answers such as the gender of pupils. Other answers relate to various categories, for example the type of school. All you can do with such answers is count them or calculate percentages. One category is no 'greater' or 'higher' in value than another. Then there are questions that are used to measure the extent to which something applies. In this case you can also calculate the mean. It is important to be constantly aware of which level of measurement applies, since this determines the way in which the data can be processed and conclusions drawn.

Secondly an instrument must meet certain **quality criteria**:

- reliability - the data must be based as little as possible on coincidence;
- validity - the instrument must measure what it is supposed to measure;
- acceptability - this relates to requirements such as clarity and honesty;
- usability - the instrument must be practicable and appropriate to the objectives.

Important objectives and important factors which can influence these objectives cannot usually be properly measured by just one question. Generally, several aspects can be attributed to a concept about which individual data can be collected. These can then be combined to measure the entire concept, for example by counting up the answers and determining the mean. Working with scales improves the chances of meeting the quality criteria of reliability and validity.

Thirdly it is advisable to **test an instrument** before it is used with a trial run.

1 The written questionnaire

1.1 Introduction

We will restrict ourselves here to the verbal interview and the written questionnaire.

The only element in interviewing which has a direct bearing on the questionnaire is the opportunity to decide during the interview what questions to ask, in what order and how to ask them. If the respondent does not give a clear or complete answer, supplementary questions can be asked. If the respondent incidentally answers the next question, the next question can be omitted. If the respondent does not understand a question, it can be asked in a different way. This facility can be seen as an advantage of the interview over the written questionnaire but if a questionnaire is really good, these problems seldom arise. However, elaborating in this way means that the benefits of using a standard approach with all respondents are lost.

One of the advantages of a written questionnaire is that more elaborate question-and-answer structures can be used. It is also possible to combine verbal and written data collection. During an interview the respondent can be presented with questions-with-answers on paper. The respondent reads the text, answers verbally and the interviewer notes down the answer. Alternatively, the respondent can be asked to answer the question on the paper itself.

We will now look mainly at written questionnaires but much of what follows also applies to interviews. In a few places specific reference is made to the interview situation.

A written questionnaire can be used to ask for many different types of information:

- general personal characteristics
- other factual information
- opinions
- attitudes
- behaviour
- experience
- appreciation
- needs
- motives
- plans
- expectations

1.2 Different types of questions

a. Open question
Open questions are rarely used in written questionnaires. Open questions are harder work both for the respondent when completing the questionnaire and for the researcher when processing the data.

Answers to open questions have to be categorised afterwards. The categories used are usually created on the basis of the answers actually given in combination with particular considerations relating to the objective or object of the research.

Good categories meet the following requirements:

- they are unambiguous;
- they are mutually exclusive;
- they are equivalent or similar;
- taken together they cover all the possible responses, although it is usually necessary to an 'other' category.

Examples:
- What do you think of the course so far?
- How do you assess the current government?

Open questions are very common in interviews. It is possible for a question to be open for the respondent, but closed for the interviewer who allocates them to categories worked out beforehand.

A written questionnaire may contain open questions because:

- the responses of the respondents cannot be anticipated, there is no opportunity to collect and categorise responses from a number of potential respondents beforehand, the question is too important to be omitted;
- the researcher has drawn up a number of response options beforehand, but does not exclude the possibility of others and adds an option: 'other, (please specify) …';
- the researcher includes an open question to motivate the respondents by giving them an opportunity to express themselves but the response is not used when processing the data.

b. Semi-open questions
If you ask several questions about a subject, each of which is aimed at a particular aspect, these questions are less open, for example:

- What do you think so far of the content of the course?
- What do you think so far of the speed of the course?

This focus on a particular aspect can also be achieved by formulating variations on sentences which the respondent must complete, for example:

- I have learned that …;
- I have learned how …;
- I have learned that it is not true that ….

c. Apparently open question
This is the type of question where clearly only one word or figure can be the right response but must be filled in by the respondents themselves. We talk about 'apparently' open questions because the answer is predetermined.

A reason for not structuring the question can be that the number of possible responses is too large or that it is more efficient to let each respondent fill it in him/herself. For example:

- What is your age?
- Where were you born?
- What is your occupation?
- How many people work in your organisation?

The place where the response is to go is often indicated. This makes the questionnaire easier to read so it is also easier to process.

d. Questions with two possible responses.
This type of question takes various forms, for example:

- a question about a fact: yes/no;
- a statement or proposition: agree/disagree;
- what is your gender? male/female;
- are you married or single? married/single.

When processing, the percentages can be determined for each response category.

e. Questions with more than two possible answers, one of which should be ticked.

For example:
What is your age?
0	25 or less
0	26-30
0	31-35
0	36-40
0	41 or more

Again, when processing, the percentages can be determined for each response category.

Comments
When dividing a quantitative variable into categories, the categories should generally be equal in range but this is not a requirement. If you can assume beforehand that few people will fall into particular categories, you can amalgamate them. Always consider what impression a question may have on a respondent. In the above example it must be clear to the respondent why people over forty should be placed in the same category and it must seem legitimate to them. In a survey relating to work, one might conclude that people over forty are not relevant to the survey - written off! Such an interpretation is clearly not conducive to participation.

f. Questions with more than two possible responses, one or more of which may be ticked.
This type of question is also referred to as a checklist and is very commonly used.

For example:

> Which national newspapers and weekly news magazines do you read regularly?
> (Tick all that apply)
> 0   none
> 0   Daily A
> 0   Evening B
> 0   News of the C
> 0   Sunday D
> 0   E Magazine
> 0   other (please specify)

Once again percentages are determined for each response category.

Comments

The response options include 'none'. If you omit this and the respondent has not ticked a response, you cannot tell if none of the categories applied or if the respondent simply omitted the question. In short, each respondent must be able to tick at least one answer. For the same reason, there is also a category 'other (please specify)'.

A format which ensures all the options are read is to provide a yes/no answer to each item as follows.

	yes	no
Daily A	0	0
Evening B	0	0
News of the C	0	0
Etc.		

g. Questions with more than two possible responses, which should be placed in sequence, or from which a certain number should be chosen.
For example, the respondent is asked to rank the options in order of priority, or to select the three most important to them.

For example:
Below is a list of restaurants serving exotic food. Please rank these types of food in order of preference by writing 1 against your favourite, 2 against your next preference and so on.

- French
- Italian
- Greek
- Spanish
- Mexican
- Chinese
- Indonesian
- Indian
- Japanese

Comments

Do not make the list too long. If it exceeds 5-8 items, it becomes too difficult for most people to rank them in a properly considered manner.

In general such rank order questions should not be used too readily for technical statistical reasons. A question of this type is easy to formulate, but processing the responses can pose problems. It would have been simpler to ask the respondent to tick their three favourite types of food. This provides less information, but the information is more reliable.

h. Questions with several possible responses arranged in a scale, one of which should be ticked.
In this type of question respondents are asked to indicate how much they agree or disagree with something, how often a certain event occurs. The response options together make up an 'ordinal scale'.

An example:
How satisfied are you with public transport in this country?
0   very dissatisfied;
0   fairly dissatisfied;
0   fairly satisfied;
0   very satisfied.

When processing, you can calculate both the percentages for each category and an overall mean.

Comments
The extremes of the response options must cover the whole range of possible responses. After 'How often …', do not begin with 'sometimes' (because there is also 'never').

You can often make do by giving the extremes and between these just indicating a scale. For example: 'entirely disagree x … x … x … x … x entirely agree'. Some researchers recommend that each possibility should be mentioned separately. The intervals must then be perceived as being equal.

Do not give too many options, the optimum number of response options is between 4 and 6. One often sees a 'middle category' in such questions, for example 'neither satisfied or dissatisfied' or 'neutral'.

Experience shows that many respondents use these categories. This is due to opting for the safe middle way or not bothering to think about it and make a choice. For this reason, we often prefer to use a number of categories like four or six without a middle category. This is called 'forced choice'.

There is also a 'no opinion', 'don't know' or 'I can't judge' category which is slightly different from a middle category. This applies if a number of respondents are unlikely to have an opinion on something or may not be able to judge. If it is omitted, the respondent will be forced to pick 'something', or not answer at all.

To prevent respondents using these categories too readily, it is best to keep them visually separate from the main response options.

For example:

    How satisfied are you with public transport in this country?

    0   very dissatisfied;
    0   fairly dissatisfied;
    0   fairly satisfied;
    0   very satisfied;
    0   I can't judge.

The response options in a scale can also be more descriptive.

For example:

    How do you feel about taking part in sport?

    0   I hate it;
    0   I don't enjoy it;
    0   I'm not particularly interested;
    0   I quite enjoy it;
    0   I enjoy it.

i. Questions with several 'items', each with several possible responses arranged in a 'scale', one of which should be ticked.

For example:

    How often do you read the newspapers and weekly magazines listed below?

	never	sometimes	regularly	often
Daily A	0	0	0	0
Evening B	0	0	0	0
News of the C	0	0	0	0
Etc.				

Comments

The form of the question as a whole is that of a matrix. With the aid of a matrix it is sometimes possible to ask several questions at once, side by side.

An example:

    Below is a list of possible things a course leader may do. Please indicate in each case how often it happens, and how often you would like it to happen?

	occurs		would like to occur	
	never	often	never	often
Give explanations	0	0	0	0
Give examples	0	0	0	0
Let the students practice	0	0	0	0
Etc.				

The advantage of efficiency can very easily become a disadvantage if a matrix gets too complicated. Respondents then easily make mistakes or are put off by a page full of checkboxes.

1.3 Use of language

Precise use of language is essential. From experience we can make the following two points:

- you learn by doing. Precise formulation is a skill that you can develop by being critical of your own writing, by asking others to comment, etc.;
- do not aim for the perfect questionnaire: it does not exist but do always learn from the errors and imperfections you discover with hindsight.

1.4 Design

Increasingly things need to be visually appealing. This improves the chances of participation and of correct responses. One problem is that not everyone finds the same style clear and attractive. If the target group is unknown or varied stick with a conservative design unless you are confident your design will appeal because it is different. In general soberness and appeal serve clarity.

With regard to the last point, the responses lend themselves to many different types of presentation: horizontal or vertical, with words, symbols or even pictures. Instructions to tick/cross, circle and shade in require different designs.

Comments

The use of lines can make things clearer, but it can also make the page too busy. If the data are to be entered by a data typist, 'circling' has advantages as the codes to be keyed in can be read off directly. If the data are going to be read in by an optical character reader, 'shade in the box' is a good technique. Make sure that the distances between response options are always the same to avoid any suggestion that the available responses are not of equal value.

1.5 Sequence and length

The sequence
In every questionnaire it is important to pay attention to the sequence in which the questions are asked. The sequence:

- must not adversely affect the interpretation of subsequent questions;
- must not suggest certain answers to subsequent questions;
- must not make the respondents skip over questions.

Common principles applied to the layout of questionnaires:

- move from general to more specific questions because if you ask specific questions first you can influence the interpretation of the general questions;
- move from factual to evaluation questions;
- move from open to closed questions.

A good strategy for focusing respondents' attention is to make them a party to what the research is about and arranging the questionnaire in sections that flow logically from this. Headers can then identify these sections.

The length

How long you make a questionnaire will depend on the capacities and willingness of the respondents and the interest they have in the research. It is a matter of judgement. In practice an hour is maximum for filling in a questionnaire and an hour and a half talking for an interview. It is not easy to say how many pages to fill or questions to ask but to give some indication, you can have up to 10 pages which are not too tightly-packed for a written questionnaire or 20 fairly specific questions in a semi-structured interview. But less is better.

1.6 Criteria for assessing the quality of the questions

There are many ways in which the quality of questions can be compromised, for example:

- the question is misunderstood or is ambiguous;
- the question is too personal;
- the question is not applicable to the respondent;
- several questions are asked at once;
- the possible responses do not cover all the possibilities;
- the possible responses are not consistent with the question;
- the intervals between points on an response scale are not equal.

The researcher can identify these problems in various ways:

- respondents give answers which are impossible;
- the answers of all the respondents together are distributed in an unexpected way, contrary to everything which is known about the subject;
- there are questions which a substantial number of the respondents have not answered;
- there is little variation in the answers to certain questions.

People also sometimes recommend that in a question with a series of items, some of them should be expressed in a negative form to counteract the tendency to give the same answer. In general, working using a mix of positively and negatively-worded items is not recommended for the following reasons:

- it is more troublesome to read and interpret negatively-worded items, so there is a greater risk of mistakes;
- most respondents do not have any such tendency as supposed;
- a positive answer to a negatively-worded question is not psychologically the same as a negative answer to a positively-worded question;
- often a scale analysis will show that the negatively-worded items do not fit in with the other items so that they do not form part of the scale.

2 The learner report

In a learner report, the respondents describe the learning experiences they have gained in an educational setting; a learner report asks about your own learning experience or the learning effects that you have experienced yourself.

The following reasons are often cited for using the learner report in evaluation research:

- the person doing the learning is central;
- the learner report gives access to some objectives which are difficult to measure, such as the affective objectives which nature and environmental education often strives for.

Forms of learner report

The learner report can be either in writing or verbally. It can take the form of:

- open questions for example, what have you learned?
- semi-structured open questions of the type:

  Complete the following sentences:

  – I have learned how …
  – I have learned that I …
  – I have learned that it is not true that …
  – I have learned that it is not true that I ….
  – questions with closed answers. In this case the respondents are given some learning effect statements and indicate the extent to which they apply to them, for example, "I have learned why environmental pollution is bad for the natural world".

The most important consideration when choosing between an open or a more closed learner report is the purpose it is intended to serve. A closed learner report is not particularly useful for tracing personal learning experiences. A closed learner report is far more suitable for comparing various groups of respondents in terms of the learning experiences they have reported.

Problems with using learner reports

The validity of the learner report can be compromised by:

- the tendency to give socially acceptable responses;
- limited skill in identifying one's own experiences;
- the lack of language skills to describe these experiences.

The problems of ordering and categorising learning experiences scarcely arise in the case of a closed learner. It also has the advantage that processing is less labour intensive. However, it does have some disadvantages. The construction of a closed learner report is, however, more laborious than that of an open one. In a closed learner report you present the respondents with learning experiences and they must then indicate whether they have experienced them or not. This can lead to respondents reporting more than they have learned.

Guidelines and suggestions for the construction of a learner report
These guidelines relate mainly to the structured learner report:

- carry out a trial run to get the questions clear;
- provide some examples of completed sentences;
- in addition to examples, you can also provide points for consideration, for example: what have you learned about woodlands? (think, for example, about the trees, the plants, light, the soil, the animals, food, waste, cycles etc);
- there are various options for sentences to be completed:
  – the respondent completes the sentence in as many ways as possible;
  – the respondent completes the sentence a limited number of times, for example the three learning experiences which made the deepest impression;
  – the respondent completes each sentence only once. In this case several sentences for completion are given to stimulate respondents to generate as many 'effect sentences' as possible.

An example of a learner report is contained in Appendix 3.

3 Knowledge test

The knowledge test is a good instrument for measuring knowledge but to be useful the question setter must consider what objectives are to be tested and which form of question is most suitable.

Forms of question

Types of open questions:
- filling in or completing. Here the respondent must complete an unfinished sentence or drawing, for example 'The stamens of a plant are used for …';
- the short answer question. Here the respondent answers with a word or a few words, numbers or simple drawings, for example 'What do you measure with a hygrometer?';
- the essay question. The respondent's answer is a coherent piece of prose in answer to a question.

Types of closed questions:
The true or false question. The respondent is presented with a statement and has to say whether it is true or false.

A hygrometer is used to measure air pressure
- 0   true
- 0   false

The proposition question. In this type of question two related statements are combined with a choice of four possible answers.

1. Water boils at 100°C
2. Every liquid has its own boiling point under normal conditions
- 0   1 and 2 are correct
- 0   1 is correct and 2 is incorrect
- 0   1 is incorrect and 2 is correct
- 0   1 and 2 are incorrect

The combination question. Here the respondent is given two columns and must link items to each other.

Match the animal group in the first column with the right animal in the second column.
- a) mammal   1) shrimp
- b) fish   2) whale
- d) mollusc   3) plaice
- e) arthropod   4) oyster

The multiple choice question. Here, a question or an unfinished statement is followed by a number of possible responses from which the respondent must choose one.

Example of a direct question:
Here is a food chain:
aphid - ladybird - woodpecker - sparrow hawk
Of which of the animals in this food chain are there **the least** number
    0   aphids
    0   ladybirds
    0   woodpeckers
    0   sparrow hawks

Example of 'omission at the end':
The stamens in a flower are used for:
    0   protecting the plant
    0   reproduction of the plant
    0   food for insects
    0   making honey

Of the above question types, the multiple choice question is the most
commonly used, partly because it takes the least time to process.

Guidelines for writing multiple choice questions

The terms used for multiple choice questions:

- the stem: the part of the question containing the question;
- the response options: the range of responses provided in the
  multiple choice question;
- the distracting statements - the incorrect answers.

The following guidelines apply to the stem:

- complete: the stem should contain all the information necessary to
  select the right option. Too little information makes it impossible
  to make a well-founded choice, while superfluous information can
  adversely affect the clarity of the question. For example,
  respondents cannot answer the following question because they do
  not know which country or time period the question relates to

Which energy sources together provide half of the total energy?
    0   natural gas, sun and wind
    0   oil and sun and wind
    0   coal and natural gas
    0   coal and nuclear power

- short: for the sake of clarity the stem should be as short as
  possible;
- short sentences: aim for sentences of no more than 12 words;
- unambiguous: the stem should not contain any words with more
  than one meaning;
- negations: where the stem is formulated in the negative, it is
  advisable to draw this to the respondent's attention by means of a
  different lay-out. Double negatives should be avoided.

Which of the energy types below is NOT 'alternative' energy?
- 0   energy from manure
- 0   nuclear energy
- 0   wind energy
- 0   solar energy

- • information: the number of data in the stem must be as low as possible. The use of subordinate clauses should be avoided.

The same rules apply to writing the response options as to the stem. The following guidelines and construction tips also apply to the responses:

- • only one of the options contains the right answer;
- • the essential differences between the options are clear;
- • the distracting statements must seem plausible for those who have not studied the material properly;
- • use learning problems and common misconceptions as distracting statements;
- • arrange the options in a logical sequence, such as ascending numerical order, alphabetical order, etc.;
- • the options must be consistent in terms of grammatical structure and general impression, including length. In addition, the options must connect to the stem in the same way;
- • avoid a situation in which two distracting statements are mutually exclusive.

Some children are walking in the woods. They pick some toadstools that they find. Is that sensible?
- 0   Yes, because toadstools are fungus and fungus is nasty.
- 0   Yes, because toadstools cause decay so that plants and trees do not get any food.
- 0   No, because toadstools cause decay so that plants and trees do get food.
- 0   No, because toadstools are far too beautiful to be picked.

The second and third options resemble each other so one of these options is sure to be the right one.
- • avoid 'highlighting' the distracting statements by the using words such as 'never' 'always' and 'all';
- • in the response options, avoid echoing concepts which were referred to in the stem.

Ladybirds were introduced into an area to control scale insects. After a number of years a 'natural balance' has developed. How has the balance come about?
- 0   the ladybirds have eaten all the scale insects
- 0   the scale insects have eaten all the ladybirds
- 0   the number of ladybirds now equals the number of scale insects
- 0   the number of ladybirds is in balance with the number of scale insects.

- avoid overlaps with other questions;
- do not use text book language or stereotyped expressions. This leads to recognition of the correct answer;
- you must be able to read an option without having first read the preceding option;
- the options must be as short as possible so that weighing them up against each other does not become too complex.

Extra requirements for use in evaluation
Questions must also meet the following requirements:

    a. relevance;
    b. objectivity;
    c. discrimination;
    d. efficiency;
    e. degree of difficulty.

(a) Relevance
A question is relevant if it measures what it is supposed to measure.

(b) Objectivity
A question is objective if different experts all correctly pick the same option. The following question fails on this criteria because it asks the respondent for an opinion:

You are better off using artificial fertiliser than farmyard manure if you want to fertilise a field of grass. Do you agree?
    0   Yes, because artificial fertiliser kills the weeds.
    0   Yes, because farmyard manure stinks.
    0   No, farmyard manure is just as good and is kinder to the environment.
    0   No, because you can only apply artificial fertiliser when it rains.

It would also fail these criteria if more than one of the options was true, or expert opinion differed about the correct answer to a question.

(c) Discrimination
A question is discriminating if it can be used to distinguish between respondents who have mastered the material from those who have not. Distracting statements should contain mistakes that respondents who have followed the course can recognise but others cannot.

Things to avoid:

- using words like 'never', 'only', 'always', or 'all';
- making the correct option noticeably longer or shorter than the incorrect ones;
- overlapping options. In the following example any respondent who knows that sand is not a source of energy, can dismiss two options straight away;

There are various sources of energy. Which of these rows contains three sources of energy?

0   natural gas, petroleum and sun
0   coal, sand and petroleum
0   wind, limestone and water
0   wind, sand and sun

- the respondents can solve the questions by looking at other questions;
- the correct answer is systematically in the same place, for example, the third option;
- not all the options follow from the stem. In that case the option which does follow is likely to be the correct one;
- a striking word recurs in the correct answer.

(d) Efficiency
A question is efficient if it provides a lot of relevant information for the respondent, but requires relatively little time for the respondent to answer.

Some features of inefficient questions are:

- a question causes reading problems for the respondents, for example when the grammatical structure of the text is complicated;
- the text is not pitched at the knowledge level of the respondents;
- superfluous information is provided.

(e) Degree of difficulty
This relates to the amount of difficulty respondents have in answering the question properly. The degree of difficulty of a question should depend on the group of respondents for whom the question is intended.

4 Skill tests

Skills are characterised by the application of knowledge to a practical situation.

Methods
To assess skills three methods can be used, the practical test, product test and the simulated practical test.

The practical test
Here a participant demonstrates their skills in a real situation. For example, if you wish to test whether a trainee primary teacher has the necessary skills to carry out fieldwork with a group of pupils, the student is sent out into the field with a group of pupils. An assessor observes whether the student is proficient in the necessary skills. The assessor uses an assessment inventory. This is a list that breaks down the skills to be tested into different elements. For each element there is a test criterion to be scored.

The product test
In the product test, the assessment is made on the basis of the result of participants' actions. For example, rather than assessing the way in which research is carried out, a researcher is assessed on their research results.

The simulated practical test
Here participants show their skills in a simulated situation such as a role play. The assessment is again based on an assessment list.

Self-assessment can be combined with the above methods to determine the skill level of the participants using a learner report.

Choosing a method
The method you select will depend on:

- the nature of the skills to be tested. A skill such as doing a somersault would obviously not be tested by having the respondents write a report on a somersault that they have done;
- how much time is available;
- the objective of the evaluation;
- the degree of validity and reliability you hope to achieve.

The following section looks at the creation of an assessment list and provides guidelines for optimising validity and reliability.

Making assessment inventories
Criteria or assessment inventories can be used to determine whether the participants are proficient in the skills.

Conditions for creating these lists are:

- the skills to be tested must be observable, and
- the skills to be tested must be capable of being broken down into smaller elements.

When selecting the criteria, a distinction is made between process criteria, product criteria and effectiveness criteria.

Process criteria: these are criteria that define the desired process, ie the various steps involved. They relate to whether the participant performs the right elements at the right time and in the right way.

Product criteria: these relate to any product of the skills used, for example a research report or work sheets for fieldwork assignments.

Effectiveness criteria: these relate to the impact of the product created by the process. Can you do anything with the research results for example?

Guidelines for improving reliability and validity
The reliability of the assessment inventory is increased if the various assessors' methods of assessment are as similar as possible. To reduce differences:

- assessors must be given clear instructions. These must go into the way in which the assessors give the test assignment to the participants, the input of the observer during the test, the duration of the various test elements and the duration of the test as a whole.
- the observers must be given clear instructions with regard to the interpretation of the various criteria. Experience shows that apparently unambiguous criteria can still be interpreted in various ways. In the case of teacher training students who are to have primary school pupils carry out a fieldwork assignment, a criterion could be: "The student supervises the pupils well". Does this mean 'intervene whenever respondents are doing something incorrectly' or 'allowing the children to work as independently as possible?'
- the criteria must be as short, but as complete, as possible.

The validity of the assessment inventory can be improved by ensuring:

- the assignments set are covered by the course of education;
- assignments are worded as clearly as possible;
- the steps in the assessment inventory are clearly brought out in the assignment;
- it is possible to complete the assignment within the given time.

5 The logbook

A logbook is a notebook or form which participants or organisers of an activity use to keep a record. A log book might be kept by the teaching staff or by the pupils themselves.

Note: Only tentative conclusions may be drawn from such a logbook, since it is a form of self-assessment. The results are mainly useful for validating other instruments.

The following logbook was used in an in-depth study into technical education carried out by an Education Research Group. The logbook provides information about the way in which technical education is designed and is divided into one general and three main categories:

1. General
2. Preparation
3. Implementation
4. Evaluation

Teachers' logbook

1. General

Name of school:

Name of teacher:

Lesson number and title:

Date of lesson:

Class:

Number of respondents present:

2. Preparation

2.1. What activities did you undertake to prepare this lesson and how much time did these take? Consider, for example, studying the content of the teaching material, copying teaching material and setting out learning materials, etc.

Activity	Time (minutes)
....	....
....	....
....	....

2.2 Did you experience any problems during the preparation of the lesson?

    0   yes

    0   no

If yes, give reasons:

....

....

3. Implementation

3.1 How much time did you spend on the following types of work in the lesson?

Teaching .... minutes

Class discussion .... minutes

Demonstrating skills (for example the use
  of a piece of equipment) .... minutes

Demonstration of a practical assignment................................... minutes

Letting respondents work on their own .... minutes

3.2 Did you devote any attention to the following activities?

1. Themes	0 yes	0 no
2. Teaching or learning concepts	0 yes	0 no
3. Teaching or learning how to apply knowledge	0 yes	0 no
4. Teaching or learning problem-solving behaviour	0 yes	0 no
5. Teaching or learning to assess workpieces	0 yes	0 no
6. Evaluation, consideration and review of chosen solutions	0 yes	0 no

3.3 What practical assignments did the students carry out? Give a brief description.

....

....

....

....

3.4 What materials were used by the students when carrying out the assignments?

....

....

....

....

3.5 What specific tools were used by the students when carrying out the assignments?
(for example, angle bending machine, polystyrene foam cutter, plastic bending form).

....
....
....
....

4. Evaluation

4.1 General impression of the lesson

useful	0	0	0	0	0	not useful
ran smoothly	0	0	0	0	0	was problematic
feasible within the time available	0	0	0	0	0	not feasible
all learning objectives achieved	0	0	0	0	0	no objectives achieved

4.2 Did you deviate from the prepared lesson structure?   0 yes   0 no

If yes, at what point, and for what reason?

....
....
....
....

4.3 During the lesson did you have any problems with the following points:

a. the structure of the learning material	0 yes	0 no
b. the application of the teaching methods	0 yes	0 no
c. grouping the students	0 yes	0 no
d. timing	0 yes	0 no
e. the teaching material and aids required	0 yes	0 no

Explanation:

....
....
....
....

4.3 How did the students behave during the lesson?

a. the students were active	0	0	0	0	0	passive
b. the students behaved independently	0	0	0	0	0	dependently
c. the students were interested in the lesson	0	0	0	0	0	not interested
d. the students were quiet	0	0	0	0	0	noisy

6 Interviews

Interviews can take several forms, open-ended, semi-structured or structured. An open-ended interview is a conversation on a given subject and the interview has no more than a number of points for consideration. For a semi-structured interview a number of open questions are prepared beforehand, which are asked in sequence. A structured interview is based on pre-defined questions and responses. Several pilot interviews should be held, the results and experiences discussed and the interview adjusted if necessary.

The following comments refer to a semi-structured interview. It may last for about an hour, in which time around 15 questions will be asked.

A list of potential respondents must be compiled and a round of interviews organised. This may be a sample of a larger population with a few reserves in case not all are willing or able to take part. The selected respondents are phoned and asked if they are willing to take part.

If the respondents are willing, an appointment is made. Three to four interviews per day is a maximum. The appointment is confirmed in writing, giving the keywords of the subjects to be discussed, how long the interview is likely to last and who the interviewer will be.

If the interview is to be recorded prior permission should obtained. Since people often have difficulty refusing a request if they are put on the spot, it is best to inform them in advance that such a request is to be made.

Writing up a recorded interview is extremely time-consuming. It is advisable to take down at least the major points directly on paper. Little information need be lost if the interviewer reads through the interview notes as soon as possible after the interview.

The interviewers must be familiar with the purpose and content of the research. If several interviewers are used, they must all interpret the questions in the same way.

A good interview will start with a short introduction to the purpose of the research, the interviewers and the organisation carrying out the research, how they were selected and how the data will be used, guaranteeing anonymity if wanted.

Although in the case of a semi-structured interview the form and sequence of the questions are laid down in advance, the interviewer will not read these out word for word. It is best to strive for as natural a conversation as possible in which the interviewer goes into what the respondent says in an interested and observant manner, occasionally asking further questions, possibly changing the sequence of the questions and so on. However the interviewer should seek to remain as close as possible to the intention of the questions and should never influence the interviewee by offering his/her own opinion. The different groups of questions should be introduced and put into a framework that the respondent understands. The interviewer should mirror

the language of the respondent and given sufficient time for thinking about the replies.

A number of respondents will have a story of their own to relate about the subject of the interview. It is often sensible to allow them to unload these at the beginning of the interview so that afterwards the interview can be completed according to plan.

The interview is closed by asking the respondents if they have anything more to add, briefly going over the agreements made and thanking them for their cooperation.

When reporting, the interviewers stick as closely as possible to the answers as they were given. Any further interpretation belongs to the processing and analysis stages. However, it is a good idea to make a note of anything which needs to be taken into account when interpreting or processing the responses, such as any disruptions which took place during the interview, the difficulty the interviewee experienced in answering certain questions, etc.

7 Observation

Here we will look at various forms of observation including the pros and cons of structured and unstructured or free observation.

Unstructured or free observation
Information collected in this manner often takes the form of descriptive or qualitative data. This form of observation can serve as preparatory work for structured observation.

Structured observation
Structured observation is concerned with exact reproduction of observations. The results of structured observations are expressed in numerical or quantitative form.

Observation systems are used which consist of a set of instructions for the observer prescribing the methods. There are different methods of recording observation:

Category systems
In category systems a particular phenomenon, event or behaviour is recorded each time it occurs.

Example A:

In this example the observer has to record the number of times a particular vehicle type passes at particular times.

Street name:

Position on street/junction:

	lorry	bus	private car	motor cycle	moped	bicycle	other
0800 - 0900							
0900 - 1000							
1000 - 1100							
1100 - 1200							
etc.							

Example B:

A well-known observation system within education theory is that of Flanders. This system has 10 categories to record the interaction between teachers and students (seven for teacher behaviour, two for student behaviour and one 'other' category). This observation system is concerned with determining the social climate in the lesson. The observer wears a headphone and every three seconds hears a beep. He or she must then record behaviour taking place. This method keeps the sequence of events intact and also provides information about the duration of events.

Rating scales
Rating scales are used to record the extent to which a phenomenon, characteristic or datum is present or, for example, how intensely a certain behaviour manifests itself. The observer marks on a scale the extent to which a certain characteristic is present. This is based on a view or impression of what has been observed.

Example C:

This example is concerned with the way in which a teacher supervises pupils.

The teacher:	—				++
gives clear examples	0	0	0	0	0
uses clear language	0	0	0	0	0
ensures that the pupils remain involved	0	0	0	0	0
has the pupils work together	0	0	0	0	0

Advantages and disadvantages of structured and unstructured observation
The advantage of structured observation is that you can verify and replicate the results.

The disadvantages are:
- the observer and object are separated: a system of rules, procedures and categories is placed between the observer and the object;
- restricted analysis: only that which can be measured and categorised plays a role in the observation. People's intentions, motives and expectations are disregarded;
- reduction of reality: meaningful behavioural entities are broken down into meaningless pieces. Often the context of the events is disregarded.

The advantages of unstructured observations are:

- complex phenomena are described and interpreted as a totality, without preconceived plans and categories;
- the free approach can be very useful in identifying categories, particularly in a descriptive research phase;
- it takes account of the fact that the results are context-related;
- the results are more comprehensible to non-researchers.

The disadvantages are:

- the difficulty to verify and replicate – it is difficult to observe without having preconceived models and categories;
- the observers use subjective categories which largely determine the results of the observation;
- the lack of unambiguous, detailed observation procedures makes replication and verification extremely difficult. It is therefore practically impossible to determine whether or not the results are reliable.

An intermediate form
An intermediate form was used in an in-depth study of technical education in schools in Holland. The observation instrument developed during this research project was intended to give as accurate a picture as possible of the activities undertaken during a technology lesson by teachers and pupils.

This instrument used structured observation based on a number of carefully selected points for consideration. The attention of the observer was directed towards certain aspects. First the various aspects of the lesson were defined:

- instruction: activities of the teacher before pupils start the assignment;
- assignments: tasks to be carried out by pupils in a given time;
- motivation: giving the reason for why something must be learned or done;
- supervision: activities of the teacher while the pupils are carrying out the assignment;
- testing: investigating whether insights, products and thinking strategies meet set criteria;
- management: activities relating to the organisation of the lesson;
- non-lesson related activities: all activities which fall outside the interaction with pupils and are not aimed at any aspects in the above categories.

Secondly, points for consideration were given for each aspect of the lesson. The following relate only to motivation and supervision.

Motivation:

- placing the assignment in a real context;
- dealing with disappointments;
- stressing the responsibility of the pupils;
- giving assignments within the context of both past and anticipated social trends;
- relating it to occupations.

Supervision:

- providing a technical explanation of the assignment to be performed to individuals or subgroups;
- demonstrating the assignment to be performed to individuals or subgroups;
- encouraging, for example, by pointing out actual ways the assignment could be carried out and the resources available;
- handling the input from pupils, for example accepting the conclusions of pupils, providing time for independent decision making, drawing out conclusions;
- discussion of any reluctance to carry out the assignments.

In addition to these, the observer was also asked to note what was happening in the lesson as a whole and was given the opportunity to report anything striking. No guidelines were given for this, so this part was unstructured. The method of registration was not prescribed either, so it was also unstructured.

Observation methods

In *participant observation* observers participate with the group they are observing. If they get closer to the group they will be able to collect data which would never come to light otherwise. On the down side, there is a risk that the observer will not function as well because they may identify with the group and interpret the information subjectively.

In *non-participant observation,* the observer fulfils the role of a neutral onlooker, observing the group from a distance such as sitting as unobtrusively as possible at the back of a school class.

Observation may be *direct* or *indirect*. In indirect observation*,* use may be made of tape, film or video recordings etc. Such tools can be useful:

- if the phenomena to be observed arise quickly or are complex;
- if they require very accurate observation, for example, slow motion or repeated replays.

Appendix II
Thirteen examples from written questionnaires

In this chapter you will find some examples of questions for pupils, teachers and directors of schools and for students doing teacher training. We provide these examples to give you an idea of the way in which you can deal with certain subjects in the form of questions and how you can put certain concepts into concrete form. It is always difficult to construct good questions. We do not pretend that these examples are some sort of panacea. You should regard them only as a source of ideas.

In reproducing the questions, we have used the same construction as far as possible, i.e.:

- aim of the research: what is to be measured;
- target group: the target group for the questions, for example, pupils in group 8 primary education or biology teachers in secondary education;
- function: the purpose for which the question was first developed including description, inventory of needs, process evaluation, effect evaluation;
- question: the question itself;
- source: the source from which the question was taken;
- results: where known, we indicate the results obtained with the question (such as percentages, means, distributions, reliability of the scales, etc).

Abbreviations and symbols used:

av = average score;
sd = standard deviation
itc = item-total correlation;
alpha = Cronbach's alpha, the most commonly used coefficient for the reliability of a scale.

It is possible that certain data do not mean very much to you. However, we consider that you can still benefit from these examples, even if you are not familiar with the statistical techniques.

Example 1

Aim of the research: to ascertain the level of basic ecological knowledge.
Target group: Pupils in upper primary education.
Function: Effect evaluation.
Question: Tick the correct answer.

We would like see what you have now found out about the natural world. It doesn't matter if you don't know very much about it. Try to answer the

questions *as best you can.* Read carefully through the questions and the possible answers. Tick what *you think is the right answer.* Always tick something. If you don't know, tick the answer that seems best to you.

Life cycles/food chains:
Lisette catches a pike in a lake. She eats the pike. There are also smaller fish living in the lake, and they eat water fleas. There are also algae growing in the lake.

What would the food chain be?

0   Lisette - pike - water fleas - smaller fish - algae
0   pike - smaller fish - water fleas - algae
0   water fleas - smaller fish - pike - Lisette
0   algae - water fleas - smaller fish - pike - Lisette

Life cycles:
No nutrients are ever wasted in nature, everything is re-used or stored.

Who or what ensures that, in the end, all natural waste is turned into nutrients for plants?

0   bacteria and fungi
0   insectivores and predators
0   herbivores
0   plants with chlorophyll

Woodland/food chains:
A food chain in a woodland might be:
birch leaf - aphid - ladybird - woodpecker - sparrow hawk

Of which of the animals in this food chain are there **the least** number?

0   aphids
0   ladybirds
0   woodpeckers
0   sparrow hawks

Natural balance:

What does 'natural balance' mean?

0   the total number of plants and animals in an area stays the same for a number of years
0   the total number of species of plants and animals in an area stays the same for a number of years
0   in a particular year there are roughly as many plants as there are animals in an area
0   the number of plants and animals of each species in an area stays roughly the same for a number of years

Natural waste:

Natural waste, which has fully decomposed, is called humus. What purpose does humus serve in the soil?

0   it offers a protective habitat for soil organisms
0   it protects the roots of plants
0   it delivers food to plants
0   it keeps soil organisms warm

Natural balance:

Ladybirds have been introduced into an area to control scale insects. After a number of years a 'natural balance' has developed. How has that balance come about?

0   the ladybirds have eaten all the scale insects
0   the scale insects have eaten all the ladybirds
0   the number of ladybirds now equals the number of scale insects
0   most of the scale insects were eaten, after which the numbers have stayed roughly the same

Food chains/woodland:

In what sequence do fallen leaves, herons, moles and earthworms form a food chain?

0   earthworms - fallen leaves - moles - herons
0   fallen leaves - earthworms - moles - herons
0   herons - moles - earthworms - fallen leaves
0   moles - herons - earthworms - fallen leaves

Natural waste:
You can make your own humus out of food and garden waste. We call it compost.

Who or what turns food and garden waste into compost?

0   bacteria, soil organisms and fungi
0   birds, mammals and insects
0   conifers and deciduous trees

Woodland:

Many European countries have pine, broad-leaved and mixed woodland. What type of woodland has the most animal species?

0   pine forests
0   broad-leaved forests
0   mixed forests

Natural balance:
In a certain area the following food chain occurs:
plants - beetles - shrews - kestrels
We know that for years there have been more kestrels than shrews present in the area.

Which explanation is correct?

0   Shrews are heavier than kestrels
0   Shrews are not the kestrels' only source of food
0   Kestrels can survive on very little food
0   Kestrels reproduce very rapidly

Life cycles/natural waste:

Every year in autumn many trees lose their leaves. Why doesn't the layer of leaves in the woods get any thicker over the years?
This is because the leaves …

0   dry out and fall apart
0   blow away after a while
0   are transformed into humus
0   get increasingly tightly-packed

Source: Karel Stokking, Lisette van Zoelen, Lisette van Aert and Ralph Young (1995). Evaluatie van nme-activitieiten, een handreiking (Evaluation of nature and environmental education activities: a helping hand), Utrecht: ISOR.

This knowledge test was developed in collaboration with the staff of the Dutch Fieldwork Foundation and a member of staff of CITO (the National Institute for Educational Measurement (CITO). It was developed for evaluating nature study weeks.

The alpha of the overall test was 0.65 in the initial measurement and 0.75 in the final measurement.

Example 2

Aim of the research: to ascertain the extent to which pupils, in their own estimation, possess the necessary skills to carry out research.
Target group: Pupils in the first stage of secondary education.
Function: Effect evaluation.
Question: Indicate how far you agree with the following statements.

  1 = disagree
  2 = agree to a limited extent
  3 = agree to a reasonable extent
  4 = agree strongly

  I can work well with others.
  I can set a research question.
  I am capable of conducting research according to a staged plan.
  I can collect data in the correct manner.
  I can produce a good research report.

Source: this scale was developed by staff of the Dutch Fieldwork Foundation in Orvelte.

This scale was measured twice, before and after a study week. Below you will find the item-total correlation and alphas. Based on the measurement before the study week, one could have concluded that all five items together form a reliable scale. In the measurement afterwards the first item (concerning collaboration) proved no longer to be consistent with the other items. It is probable that the pupils gained more collaborative experience doing the research project (positive and negative) and this has changed their estimation of their skills. The analyses were therefore run again, omitting the first item.

The item-total correlation below relate to the analyses of all five items. The alpha results are given for both the five and four-item results.

	itc	
	before	after
I can work well with others	0.28	0.01
I can set a survey question	0.54	0.48
I am capable of conducting research according to a stepped plan	0.56	0.73
I can collect data in the correct manner	0.66	0.57
I can make a good research report	0.61	0.55
Alpha of all five items	0.76	0.69
Alpha of items 2-4	0.78	0.79

If you intend to make comparisons, as in this case with two sets of measurements, it is always desirable to use such a method to verify the reliability of each measurement separately.

Example 3

Aim of the research: Ascertain the extent of nature and environmentally friendly attitudes.
Target group: Pupils in upper primary education.
Function: Effect evaluation.
Question: Tick the correct answer.

Willingness to take action

	strongly disagree	agree slightly	agree largely	agree completely
I would like to take part in a demonstration against the felling of forests.	0	0	0	0
I would be happy to collect signatures for the protection of endangered species.	0	0	0	0
I would be happy to take action for Greenpeace.	0	0	0	0
I would like to be a member of a natural history club.	0	0	0	0
I would like to write something in the school newspaper about protected species.	0	0	0	0

Source: Karel Stokking, Lisette van Zoelen, Lisette van Aert and Ralph Young (1995). *Evaluatie van nme-activitieiten, een handreiking*, Utrecht: ISOR (see above).

Example 4

Aim of the research: Ascertain the extent of nature and environmentally friendly attitude.
Target group: Pupils in the first stage of secondary education.
Function: Effect evaluation.
Question: Look at the sentences below about nature. Put a cross in the relevant place after each sentence to show whether you disagree entirely, disagree, agree or agree entirely.

> 1 = disagree entirely
> 2 = disagree
> 3 = agree
> 4 = agree entirely
>
> I like going on field trips with the forestry warden.
> I like going for walks in the country at the weekends / during the holidays.
> I like reading books about nature.
> I like watching nature programmes on the TV.
> I would like to take part in a demonstration against the felling of forests.
> I would be happy to collect signatures for the protection of endangered species.
> I would like to be/already am a member of a natural history club.
> I would like to write something in the school newspaper about, say, a Greenpeace campaign.

The alpha of the scale as a whole was 0.78 in the initial measurement and 0.76 in the final measurement.

Source: staff of the Dutch Fieldwork Foundation in Orvelte developed this scale.

The scale can be seen as an extension of the scale described above in which the attitude to nature and the attitude to the environment can be measured separately.

Below are the item-total correlation for the separate sub-scales (each of four items) and the results over the entire scale (8 items).

I like going on field trips with the forestry warden.		0.66	0.49
I like going for walks in the country at the weekends or during the holidays.		0.64	0.59
I like reading books about nature.		0.51	0.60
I like watching nature programmes on the TV.		0.27	0.21
I would like to take part in a demonstration against the felling of forests.		0.47	0.68
I would be happy to collect signatures for the protection of endangered species.		0.44	0.65
I would like to be/already am a member of a nature club.		0.57	0.78
I would like to write something in the school newspaper about, say, a Greenpeace campaign.		0.21	0.22
alpha	.70	.64	.80

The data show that the second group of items about the environment is a weaker scale than the first group about nature. In both cases the fourth item in the group fits slightly less well. The overall scale is still reliable. The fourth and eighth items still have a fairly low item-total correlation in the scale.

Based on these results, attitudes to nature and the environment could be handled separately. However, there is also much to be said for taking them together in a single attitude to nature and the environment.

Example 5

Aim of the research: Ascertain the level of appreciation of nature.
Target group: Pupils in upper primary education.
Function: Effect evaluation.
Question: Tick the correct answer.

Below there is series of sentences. For each sentence put a tick to indicate whether you strongly disagree, agree slightly, agree largely, or agree completely. If you don't know, miss out that sentence and go on to the next one.

Intriguing nature:
     I think it's exciting that if you look carefully you can see a lot of little creatures in the lifeless soil.
     It is good fun to look for mouse holes.
     It is good fun to look in birds' nests.
     I think it's exciting to follow the track to an anthill.
     I think it's really interesting that you can look at all the things that are in the soil.

(alpha before: 0.72 and after: 0.77)

     Aesthetic nature:
     The colours of the autumn leaves are beautiful.
     The woodland smells wonderful after it has rained.
     It is great that there are so many different animals.
     Flowers are very pretty.
     It is fun to collect stones.

(alpha before: 0.72 and after: 0.71)

Source: Karel Stokking, Lisette van Zoelen, Lisette van Aert en Ralph Young (1995). *Evalueren van nme-activiteiten, een handreiking*. Utrecht: ISOR (see above).

Example 6

Aim of the research: Gauge the feelings of pupils about nature and the environment.
Target group: Pupils in upper primary education.
Function: Description of the status quo.
Question: Please indicate below whether for you these statements are 'not true', 'slightly true', 'fairly true' or 'entirely true.'

	entirely untrue	slightly true	fairly true	entirely true
1. I am afraid that in the future there will be more and more natural disasters	0	0	0	0
2. As long as only animals suffer from environmental pollution, I'm not too bothered	0	0	0	0
3. I am afraid of having to live in a world without trees and animals	0	0	0	0
4. I feel sad when I hear that children have got sick because of environmental pollution	0	0	0	0
5. I am not afraid of natural disasters in the future	0	0	0	0
6. I feel sad when I think about ailing trees	0	0	0	0
7. It makes me cross that industry pollutes the environment	0	0	0	0
8. It annoys me that many adults do not care for the environment	0	0	0	0
9. I believe it is worthwhile if I do something for the environment	0	0	0	0
10. I get really cross when I hear that a shipping accident has caused large oil slicks into the sea	0	0	0	0
11. I believe that people will find a way of saving the environment just in time	0	0	0	0
12. Anything I could do for the environment would be pointless	0	0	0	0
13. Environmental pollution is nowhere near as bad as people make out	0	0	0	0
14. There is so much talk about environmental pollution that I don't listen any more	0	0	0	0
15. I sometimes worry a lot about environmental pollution	0	0	0	0
16. I am afraid that problems in the third world will never be solved	0	0	0	0

Source: Original source: Szagun, G. & V.I. Pavlov (1993). Umweltbewusstsein bei deutschen und russisschen Jugendlichen: ein interkultureller vergleich. In: G. Eulefeld, *Studien zur Umwelterzeihung. Band 2.* Kiel: Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel.

The items were translated into Dutch by Bastings, M.A.S. & K.M. Stokking (1996). In *Leerlingen in relatie tot natuur en milieu. Een onderzoek naar kennis, houding en gedrag van leerlingen in het basis - en voortgezet onderwijs ten aanzien van natuur en milieu* (Pupils in relation to nature and the environment. A study of knowledge, attitudes and behaviour of pupils in primary and secondary education with regard to nature and the environment) Utrecht: University Utrecht, ISOR.

This scale was used with pupils in upper primary education and in lower secondary education.

Items 2, 5, 7, 11 and 13 are worded in the positive, rather than in the negative. For the sake of the scale analysis these item scores are converted (score 1 becomes 3, and vice versa). This is explained further below.

Twelve items (ie, all except 11, 12, 14 and 16) form a reliable scale with an alpha of 0.76 for all (945) pupils together.

The scale also proved sufficiently reliable for each separate group: in primary education the alpha is 0.68 and 0.70 and in secondary education, 0.75, 0.73 and 0.82 for different age groups.

Example 7

Aim of the research: Determine the extent to which pupils are inclined to give socially acceptable answers.
Target group: Pupils in primary and secondary education.
Function: Effect evaluation. (Questions put by educational establishments often relate to preferences and attitude. Particularly when people are reporting on themselves, we must bear in mind that their reports can be sensitive to what is socially acceptable. People may give the socially acceptable answer rather than their true opinion. To investigate this, a scale can be included which measures the tendency to give socially acceptable answers. The results of this scale can be used to correct the answers to other questions where necessary.)

Question: Read the sentences below and indicate whether you entirely disagree, agree slightly, largely agree or entirely agree. (People have different views about nature and the environment. This also makes them behave differently. The same applies to many other areas of life. Everyone is different: one person may be quick to lose their temper and be grouchy, another person always stays calm and good-tempered. What type of person are you?)

1. I am never jealous.
2. Sometimes I make sure I get my own way.
3. If I have to wait a long time I can sometimes lose patience.
4. I lie a bit sometimes.
5. I am always honest.
6. I always stick to the rules.
7. I always do what I'm asked straight away.
8. I am sometimes unkind.
9. I am never impolite.

Source: Original source: Grift, W. van de, Vos, W. de & J. Meloen (1991). *Antisemitisme en autoritarisme onder scholieren* (Anti-Semitism and authoritarianism among school pupils). De Lier: Academisch Boeken Centrum. (In collaboration with E. Kliphuis). This survey was conducted on 867 school children (lower vocational education, lower general secondary education, higher general secondary education, pre-university education). At the time of the survey most of the pupils concerned were between 15 and 18 years old.

The items have been used by, among others: Stokking, K., Van Zoelen, L., Van Aert, L. & R. Young (1995). *Evalueren van nme-activiteiten. Een handreiking* (Evaluation of nature and environmental education activities: a helping hand), Utrecht: Universiteit Utrecht, ISOR.

This research concerned a study of the results of fieldwork weeks for pupils in upper primary education, organised by SVN (the Dutch Fieldwork Foundation) Apeldoorn.

Results as obtained in the survey by Stokking, Van Zoelen, Van Aert and Young

    1 = entirely disagree
    2 = agree slightly
    3 = largely agree
    4 = entirely agree

	av. (1-4)	sd	itc
1. I am never jealous.	2.57	0,93	0.23
2. Sometimes I make sure I get my own way.	2.31	0.99	0.16
3. If I have to wait a long time I can sometimes lose patience.	2.33	1.02	0.33
4. I lie a bit sometimes.	2.43	0.96	0.24
5. I am always honest.	2.39	0.92	0.48
6. I always stick to the rules.	2.58	0.85	0.40
7. I always do what I'm asked straight away.	2.43	0.84	0.43
8. I am sometimes unkind.	2.36	0.85	0.32
9. I am never impolite.	2.62	0.98	0.21

Further information about converting item scores

If you want to calculate the mean or if you want to carry out scale analysis, the response categories must have the same scoring system for all the items. If that is not the case, for example because some items are formulated in a positive and some in a negative way, the 'odd' items have to be coded 'the other way round'. For this purpose we usually take the smallest group of items or the group which is constructed negatively. In this example, it comes down to the same thing: the five items 1, 5, 6, 7, and 9 express socially acceptable behaviour, and the remaining four items 2, 3, 4, and 8 are, by contrast negatively formulated. To code the other way round, we convert the scores: 1 becomes 4, 2 becomes 3; 3 becomes 2 and 4 becomes 1. Now all the items point in the same direction.

The above scale was used with pupils in upper primary education (n=92). The scores for items 2, 3, 4 and 8 have been converted.

In the survey of Stokking et al, the alpha of this scale was 0.62. This is on the low side. The mean on the overall scale for all pupils was 2.45.

In the research by Van de Grift et al six response options were given, instead of four, namely: agree entirely; largely agree; agree slightly; disagree slightly; largely disagree; disagree entirely. The alpha in this case was 0.76.

By calculating the correlation between this scale for social acceptability and other scales or questions one can determine how much respondents have actually given their true opinion or rather answered as they felt they ought to answer. Thus in the research by SNV Apeldoorn, the correlation was calculated between social acceptability on the one hand and, among other things, willingness to take action and make sacrifices on the other. The correlation between social acceptability and how much they enjoyed the fieldwork week was also calculated. The analysis showed that these correlation were not significantly greater than 0, in other words: the pupils were not influenced in their responses by a tendency to give socially acceptable answers.

Example 8

Aim of the research: Determine attitude and behaviour of head teachers and teachers in relation to nature and the environment.
Target group: head teachers and teachers in primary and secondary education.
Function: Establish current state of affairs.
Question: To what extent do you agree with the statements below in terms of your own behaviour or attitude in relation to nature and the environment?

> 1 = disagree
> 2 = agree slightly
> 3 = agree largely
> 4 = agree entirely

1. I think nature is wonderful; I like walking in the country.
2. I often read newspaper articles about nature and the environment.
3. I often read nature and/or environmental magazines.
4. I regularly watch/listen to nature and environmental programmes on TV and radio.
5. I talk to other people about how they can contribute to a cleaner environment.
6. I try to convince other people to behave in a way that is better for nature and the environment.
7. I comment if other people behave in a way that is environmentally unfriendly.
8. I regularly go out to watch wildlife.
9. On walks, I collect stones, feathers, shells, seeds or fruits.
10. I like to identify plants, fossils etc.

Sources: This scale was used in a national evaluation study on the introduction of nature and environmental education into various sectors of education. The following sub-reports have been published on the subject: Mellema, R (1994). *Stand van zaken nme in het primair onderwijs. Najaar 1993* (State of affairs in nature and environmental education in primary education, Autumn 1993). Utrecht: Universiteit Utrecht, ISOR.

Stokking, K., Mellema, R. & R. Young. (1995). *De invoering van nme in het primair onderwijs. Stand van zaken najaar 1994 (*The introduction of nature and environmental education into primary education. State of affairs, Autumn 1994). Utrecht: Universiteit Utrecht, ISOR.
Young, R. (1994). *Stand van zaken nme in het Voortgezet Onderwijs. Najaar 1993* (State of affairs nature and environmental education in secondary education, Autumn 1993). Utrecht: Universiteit Utrecht, ISOR.
Young, R. (1995). *Stand van zaken nme in het Voortgezet Onderwijs, Najaar 1994.* (State of affairs nature and environmental education in secondary education, Autumn 1994). Utrecht: Universiteit Utrecht, ISOR.

A final report has also been published:
Stokking, K., R. Young, L. van Zoelen, F. Leenders, & M. Bastings. (1996). *Tussen vraag en aanbod. Eindrapport van het evaluatieonderzoek naar de invoering van natuur- en milieu-educatie in het onderwijs (1991-1995).* (Between demand and supply. Final report on the evaluation research into the introduction of nature and environmental education into schools (1991-1995). Utrecht: ISOR.

In the survey, head teachers and teachers in primary and secondary education were asked a number of questions about their own behaviour/attitude in relation to nature and the environment. The underlying question was, 'do head teachers and teachers who have a more positive and/or active attitude do more in the way of nature and environmental education?'. The table below shows the percentages of the head teachers and teachers who largely or entirely agreed with the statements:

	primary education		secondary education	
	heads	teachers	heads	teachers
1. I think nature is wonderful; I like to go walking in the great outdoors.	87	90	97	95
2. I often read newspaper articles concerning nature and the environment.	80	79	92	88
3. I often read natural science and/or environmental magazines.	66	61	71	52
4. I regularly watch/listen to nature and environment programmes on TV and radio.	62	70	62	62
5. I discuss with others how they can contribute to a cleaner environment.	53	52	59	46
6. I try to convince other people to behave in a way which is better for nature and the environment.	57	57	52	43
7. I comment if other people behave in a way which is environmentally-unfriendly.	54	53	52	47
8. I regularly go out to watch wildlife.	27	28	40	32
9. On walks, I collect stones, feathers, shells, seeds or fruits.	22	31	31	26
10. I like to identify plants, fossils, etc.	18	18	31	21

The table shows that there is little variation between the four groups: 90-95% like being in the countryside, 80-90% read newspaper articles about it, around 60% read magazines about nature and the environment and watch and listen to programmes about it, while a good 50% are active advocates for nature and the environment in their communication with others.

The items in all the respondent groups together did not prove to form an equally good scale. A good scale was obtained in the case of teachers in primary education in 1994 (alpha 0.85). Further analysis showed that items 2-4 always formed a good scale (re. awareness of nature and the environment) as did items 5-7 (re. active advocacy) and items 1 and 8-10 (re. leanings towards the natural environment). Analysis also showed that the level of interest among teachers in nature and the environment and in discussing it with others, corresponds positively with the amount of nature and environmental education activities they carry out.

Example 9

Aim of the research: Investigate activities of head teachers in respect of nature and environmental education.
Target group: head teachers in primary and secondary education.
Function: Establish current state of affairs.
Question: to what extent do you develop the following activities/forms of support for further development of nature and environmental education in your school?

> not at all
> to a small extent
> to a reasonable extent
> to a large extent

1. I urge teachers to pay attention to nature and environmental education.
2. I see to it that every teacher gains experience in nature and environmental education.
3. I try to ensure that as many teachers as possible are involved in the introduction of nature and environmental education at school level.
4. I allocate tasks to different teachers when a new nature and environmental education activity starts up.
5. I stimulate initiatives by teachers to make contacts in the area of nature and environmental education with other schools or nature and environmental education organisations.
6. I provide information for the organisational and financial aspects of projects.
7. I read incoming information about nature and environmental education myself and select what to distribute to the interested parties.
8. I pass on information about nature and environmental education to teachers to benefit teaching in various subjects.
9. I respond enthusiastically to initiatives by teachers in the area of nature and environmental education.
10. I try to extend initiatives in nature and environmental education to include the whole team.
11. I try to get nature and environmental education accepted by the whole team.
12. I provide support for nature and environmental education initiatives, including financial support.
13. I try to organise matters to accommodate nature and environmental education (adjusting the timetable, etc.).

Sources: Stokking, K., Mellema, R. & R. Young (1995). *De invoering van nme in het primair onderwijs. Stand van zaken najaar 1994.* Utrecht: Universiteit Utrecht, ISOR.
Young, R. (1995). *Stand van zaken nme in het Voortgezet Onderwijs. Najaar 1994.* Utrecht: Universiteit Utrecht, ISOR.
Stokking, K., R. Young, L. van Zoelen, F. Leenders, & M. Bastings (1996). *Tussen vraag en aanbod. Eindrapport van het evaluatieonderzoek naar de invoering van natuur- en milieu-educatie in het onderwijs (1991-1995).* Utrecht: ISOR.

The percentages below indicate how many head teachers (n = 154) said that they develop activities/ forms of support to a reasonable or large extent.

1. I urge teachers to pay attention to nature and environmental education. **42%**
2. I see to it that every teacher gains experience in nature and environmental education. **12%**
3. I try to ensure that as many teachers as possible are involved in the introduction of nature and environmental education at school level. **31%**
4. I allocate tasks [to different teachers] when a new nature and environmental education activity starts up. **30%**
5. I stimulate initiatives by teachers to make contacts in the area of nature and environmental education with other schools or nature and environmental education establishments. **46%**
6. I provide information for the organisational and financial aspects of projects. **45%**
7. I read incoming information about nature and environmental education myself and select what to distribute to the interested parties. **54%**
8. I pass on information about nature and environmental education to teachers to benefit teaching in various subjects. **83%**
9. I respond enthusiastically to initiatives by teachers in the area of nature and environmental education. **83%**
10. I try to extend initiatives in nature and environmental education to include the whole team. **46%**
11. I try to get measures in the area of nature and environmental education accepted by the whole team. **53%**
12. I provide support for nature and environmental education initiatives, including financial support. **57%**
13. In organisational terms I try to arrange matters to make nature and environmental education operate more smoothly (adjusting the timetable etc). **48%**

The response options were coded as follows:

not at all = 1
to a small extent = 2
to a reasonable extent = 3
to a large extent = 4

The extent to which head teachers develop the above activities for further development of nature and environmental education proves to be consistent (alpha in primary education 0.86, in secondary education 0.87). This

indicates that the head teachers vary in the extent to which they develop support activities, and not so much in the choice of the precise form of support which they offer to teachers.

Example 10

Aim of the research: Determine what working methods teachers in nature and environmental education use.
Target group: Teachers in primary and secondary education.
Function: Establish current state of affairs.
Question: To what extent do you use the following working methods to deal with subjects relating to nature and the environment?

>   not at all
>   to a small extent
>   to a reasonable extent
>   to a large extent

Primary education:
>   1. discussion of topical events concerning nature and the environment
>   2. getting pupils to make observations of nature and the environment
>   3. group discussions with or between the pupils
>   4. getting pupils to conduct research in the vicinity of the school
>   5. getting pupils to conduct experiments
>   6. getting pupils to do fieldwork
>   7. getting pupils to role play

Secondary education:
>   1. discussion of topical events concerning nature and the environment
>   2. getting pupils to work in groups
>   3. getting pupils to conduct experiments
>   4. getting pupils to conduct research
>   5. getting pupils to hold group discussions
>   6. getting pupils to do outdoor work around the school
>   7. offering pupils neighbourhood studies
>   8. getting pupils to role play

Sources: This scale was used in a national evaluation study on the introduction of nature and environmental education into various sectors of education. The following sub-reports have been published on the subject:
Stokking, K., Mellema, R. & R. Young (1995). *De invoering van nme in het primair onderwijs. Stand van zaken najaar 1994 (*The introduction of nature and environmental education into primary education. State of affairs, Autumn 1994). Utrecht: University Utrecht, ISOR.
Young, R. (1994). *Stand van zaken nme in het Voortgezet Onderwijs. Najaar 1993.* (State of affairs nature and environmental education in secondary education, Autumn 1993). Utrecht: University Utrecht, ISOR.
Young, R (1995). *Stand van zaken nme in het Voortgezet Onderwijs. Najaar 1994.* (State of affairs nature and environmental education in secondary education, Autumn 1994). Utrecht: University Utrecht, ISOR.

Mellema, R. (1994). *Stand van zaken nme in het primair onderwijs. Najaar 1993* (State of affairs in nature and environmental education in primary education, Autumn 1993). Utrecht: University Utrecht, ISOR.

A covering final report has also been published:
Stokking, K., R. Young, L. van Zoelen, F. Leenders, & M. Bastings (1996). *Tussen vraag en aanbod. Eindrapport van het evaluatieonderzoek naar de invoering van natuur- en milieu-educatie in het onderwijs (1991-1995)* (Between demand and supply. Final report on the evaluation research into the introduction of environmental education into schools (1991-1995). Utrecht: ISOR.

In the survey primary school teachers were asked which working methods they used. The table below shows the percentages of teachers who say that they use the methods referred to regularly or often (*means: this item was not included in the relevant questionnaire).

Primary education:

	1993	1994
1. discussion of topical events concerning nature and the environment	73	61
2. getting pupils to make observations of nature and the environment	35	34
3. group discussions with or between the pupils	25	22
4. getting pupils to conduct research in the vicinity of the school	24	19
5. getting pupils to conduct experiments	17	16
6. getting pupils to do fieldwork	11	15
7. getting pupils to role play	9	13.

Secondary education:

	1993	1994
1. discussion of topical events concerning nature and the environment	*	43
2. getting pupils to work in groups	39	39
3. getting pupils to conduct experiments	24	23
4. getting pupils to conduct research	26	22
5. getting pupils to hold group discussions	24	23
6. getting pupils to do outdoor work around the school	10	*
7. offering pupils neighbourhood studies	*	21
8. getting pupils to role play	3	3

Clearly nature and environmental education in primary education usually take the form of group discussions. A minority of around 30% of teachers take pupils out for outdoor activities. There is very little change between 1993 and 1994 except for a decrease in group discussion.

In secondary education we see a similar picture, albeit there are clearly fewer discussions with pupils than in primary education.

Thus one can draw the conclusion that teaching methods considered important for nature and environmental education such as discussion and

role-playing are used infrequently. Equally, methods which are important for developing values in relation to nature, such as working outside and neighbourhood studies are also used infrequently.

Example 11

Aim of the research: Investigate motivation for and objectives of participation in study weeks among primary school teachers and school managers.
Function: Effect evaluation. The items are included in the questionnaires for both the teachers and management. In addition, items for pre-activity and post-activity measurements are included.
Question: The school managers were asked:
To what extent do the motives and objectives below play a role in the participation of your school in this study week?

The teachers were asked:
- Initial measurement: To what extent did the motives and objectives below play a role in your own participation in this study week?
- Final measurement: What is your general impression of the extent to which the objectives below were achieved by the study week?
- Final measurement: On which objectives would you place the emphasis the next time?

not at all
to a small extent
to a reasonable extent
to a large extent

1. Pupils get to know each other better.
2. The mutual bonds between pupils are reinforced.
3. We have an enjoyable week together.
4. Pupils learn to work together better.
5. Pupils learn to help each other.
6. It promotes the independence of the pupils.
7. Pupils experience the pleasure of the natural world.
8. Pupils learn about nature in a different way.
9. Pupils learn different things to what they learn at school.
10. Pupils learn more about nature.
11. Pupils gain better insight into the natural world.
12. Pupils change their attitude to nature.
13. Pupils change their behaviour in relation to the natural world.
14. Pupils have positive experiences of the natural world.
15. Pupils begin to appreciate nature better.
16. Pupils learn research skills.

Sources: Stokking, K., Van Zoelen, L., Van Aert, L. & R. Young (1995). *Evalueren van nme-activiteiten. Een handreiking* (Evaluation of nature and environmental education activities: a helping hand) Utrecht: Universiteit Utrecht, ISOR.

The list consists of two separate scales, namely:
- items 1-8: objectives aimed at cooperation between pupils, alpha = 0.73;
- items 9-16: objectives aimed at pupils and the natural world, alpha = 0.79.

Example 12

Aim of the research: Determine appreciation of supervision received.
Target group: primary school pupils who have completed a fieldwork week.
Function: Process evaluation.
Question: What do you think of the supervision you received from the course tutor?

1 = poor
2 = unsatisfactory
3 = satisfactory
4 = good

His/her knowledge of the subject	1	2	3	4
Method of supervision	1	2	3	4
Enthusiasm	1	2	3	4
Organisation	1	2	3	4
Teaching skills	1	2	3	4

Source: staff of the Dutch Fieldwork Foundation in Orvelte developed this scale.

The item-total correlation for 24 primary pupils following a fieldwork course were 0.42, 0.55, 0.70 and 0.41 respectively and the alpha was 0.73. The pupils' appreciation of the supervision may therefore be determined in a sufficiently reliable manner by taking the average of their scores on these four items.

Example 13

Aim of the research: Determine appreciation of supervision received.
Target group: primary teacher training students who have completed a fieldwork week.
Function: Process evaluation.
Question: What do you think of the supervision you received from the course tutor?

a. His/her knowledge of the subject
0 poor   0 unsatisfactory   0 satisfactory   0 good
b. Extent and frequency of supervision
0 poor   0 unsatisfactory   0 satisfactory   0 good
c. Enthusiasm
0 poor   0 unsatisfactory   0 satisfactory   0 good
d. Organisation
0 poor   0 unsatisfactory   0 satisfactory   0 good
e. Teaching skills
0 poor   0 unsatisfactory   0 satisfactory   0 good

Source: staff of the Dutch Fieldwork Foundation in Orvelte developed this scale.

The scale analysis produced the following results:

	item-total correlation	alpha following removal of item b
item a	0.47	0.56
item b	-0.01	0.75
item c	0.45	0.57
item d	0.47	0.55
item e	0.66	0.46

In theory, based on the alpha, one could say that, while it is a weak scale, the average of all five items in any case gives a reasonable picture of the appreciation of the supervision (better than the answers to individual items). However, the item-total correlation of item b is around 0. Therefore this item, in this target group, is not consistent with the other items and thus does not measure the same thing as the other items. If we remove this item, the alpha rises to 0.75. The average of items a, c, d and e is now a good (reliable) measure of the students' appreciation of the supervision.

The difference in the results of the scale analysis compared with the previous example (for primary pupils) may have to do with the fact that trainee teachers and pupils have a different understanding of 'supervision'. Furthermore, in this case this item was differently formulated, with slight duplication: 'extent and frequency'. This may have caused some confusion.

Appendix III
An example of a learner report

This learner report was designed for the evaluation of a nature study course for pupils in upper primary education. A learner report for younger pupils in secondary education could be practically the same, since the questions in a learner report should be kept as short and simple as possible.

NB: Due to printing constraints the layout below differs from the original.

Learner report for pupils, to be completed at the end of the course.
On the following pages you can describe, in your own words, what you learned during the course and what you thought of it. Because it is not easy to know what to write, we will give you some help as you go along.

Knowledge
First we would like to know what you think you have learned about woodland and forests.
You can write what you have learned using different kinds of sentences such as:

I have learned that..

I have learned how ....

I have learned why ....

Below are some sentences you can complete by saying what you have learned. One sentence is enough. If you know more, that is fine too. Only write down things that you really have learned. Do not write down things which you knew already. Use different sentences to the ones shown above. [*Hand-written examples were given In the learner log used.*]

What have you learned about woodlands and forests?
(Think, for example, about the trees, the plants, the light, the soil, the animals, food, waste, life cycles, etc.)

I have learned that..

I have learned that ....

I have learned how ....

I have learned how ....

I have learned why ....

I have learned why ....

Experience

Now we would like to know what you thought of the study week and why. For example:

[*Hand-written examples were given in the learner log used.*]

Taking part in a study week was ...., because ....
or:
Taking part in a study week was ...., because ....

Once again, fill in the sentences below. Only write about things you experienced yourself. If you only complete one or two sentences, that is fine. Write *different* sentences to the ones above.

How did you enjoy being on the study course?
(Think, for example, about working in the woods, playing in the woods, staying in the study-week house, the project leader, the other children, etc.)

Taking part in a study week was ...., because ....

Taking part in a study week was ...., because ....

Taking part in a study week was ...., because ....

Taking part in a study week was ...., because ....

Liking

Perhaps you now like being in the countryside better than before the study week. Then again, perhaps you liked it better and thought it was more beautiful before the study week. Perhaps you now think that nature is very valuable. Or perhaps you have decided that the natural world is not actually as special as you had always thought.

The following sentences are about this. Again, fill them in. Only write about things you really feel. If you can only complete one or two sentences, that is fine.

For example:
[*Hand-written examples were given here in the learner log used.*]

I think the natural world is ....
or
I think the natural world is ....

What do you think of nature now? (the woods, the trees and the plants, animals and water?)

I think the natural world is ...., because ....

I think the natural world is ...., because ....

I have discovered that ....

I have noticed that...

Attitudes

Often when people have learned or experienced something new they decide to do something. We can imagine that you have decided to do something now you have finished the study week. That could be something to do with nature studies, or it could be something else. Write down below what you intend to do. Only write down what you really mean to do. If you can only complete one or two sentences, that is fine. In any case, try to write something that has to do with nature.

Plans which have to do with nature:

From now on I will try to ………………….

From now on I will try to ………………….

Other plans:

From now on I will try to ………………….

From now on I will try to ………………….

Further Reading

Bennett D B (1984) *Evaluating environmental education in schools: a practical guide for teachers.* Unesco-UNEP International Environmental Education Programme.

CASME (1994) *A skills framework for assessment in science education*. CASME, University of Natal, Durban.

Carr, W. and Kemmis, S. (1984) *Curriculum evaluation: Philosophical and procedural dilemmas.* Deakin University, Melbourne.

Carr, W. and Kemmis, S. (1986) *Becoming critical*. Falmer Press, London.

Dilley, L. (1993) *An evaluation of educational media initiatives in South Africa*. SACHED Trust and the Argus group, Johannesburg.

Dorion C (1993). *Planning and evaluation of environmental education: primary.* WWF-UK/CEE.

Dorion C (1993). *Planning and evaluation of environmental education: secondary.* WWF-UK/CEE.

Feuerstein, M. (1986) *Partners in evaluation: Evaluating development and community programmes with participants.* MacMillan, London.

Harris G and Blackwell C (eds) (1996). *Environmental issues in education*. Arena, Ashgate Publishing Limited.

Hungerford H and Volk T (1990). *Changing learner behaviour through Environmental Education*. Journal of Environmental Education, 21(3): 8-21.

IUCN (1999) *Learning in Protected Areas- How to Assess Quality, Report of a European Conference Austria and Hungary- September 1998.* European Committee for Environmental Education ECEE, IUCN Commission on Education and Communication.

Jacobson S K (1991*). Evaluation model for developing, implementing and assessing conservation programmes: examples from Belize and Costa Rica*. Environmental Management, 15(2): 143-150.

Jacobson S K (1995). *Conserving wildlife - international education and communication approaches.* Columbia University Press, New York. (Lists other references).

Jacobson S K (1999). *Communication skills for conservation professionals.* Island Press, Washington DC.

Leal Filho W and Taylor C (1995). *A framework for the monitoring and evaluation of environmental education projects in industrialised and developing countries.* ERTCEE, Bradford.

Morvaridi D (1994). *Sustainable development and project appraisal.* In Weiss J (ed) *The economics of project appraisal and the environment.* Edward Elgar Publications, London.

Nowak P (1984). *Direct evaluation: a management tool for program justification, evaluation and modification.* Journal of Environmental Education, 15(4): 27-31.

O'Donoghue, R.B. (1986) Environmental education and evaluation: An eleventh hour reconciliation. *Southern African Journal of Environmental Education,* 3, 18-21.

O'Donoghue, R.B. and Taylor, J. (1988b) *A Handbook of evaluation techniques for environmental interpretation.* Natal Parks Board, Pietermaritzburg.

Padua S (1994). *Conservation awareness through an environmental education programme in the Atlantic Forest of Brazil.* Environmental Conservation, 21(2): 145-151.

Padua S, Tabanez M and Souza M G (1997). *How can nature trails be more effective?* In *Environmental Education for the Next Generation.* Selected papers from the 25th Annual Conference of the North American Environmental Education Association, San Francisco, 1996: 263-267.

Ramashia R and Rankin S (1995).*Managing evaluation: a guide for NGO leadership.* Pact South Africa, Braamfontein.

Salant P and Dillman D A (1994). *How to conduct your own survey.* John Wiley and Sons, New York.

Shufflebeam D et al (1971). *Educational evaluation and decision making.* Hasca IL: FE Peacock (Pub).

Stokking K (1994). *Implementation of environmental education: an evaluation of a national project.* Natuur-en Milieu Educatie, Utrecht (mimeo).

Stokking K et al (1995). *Evaluating activities in environmental education: a helping hand.* University of Utrecht, Utrecht.

Van Pelt J (1990). *Project appraisal and sustainability: methodological challenges.* In Project Appraisal, Vol 5, No 3.

Wholey J S et al (19??). *Handbook of practical program evaluation.* Jossey-Bass, San Francisco.