

Newby, T. (1992). Training Evaluation Handbook (pp. 96-119). San Diego, CA: Pfeiffer.

7

The interview

Interview is an appropriate method to adopt for a number of evaluation purposes:

- Exploration of learners' responses to particular training designs or learning methods.
- Assessing the extent to which training aims and content are perceived to be relevant to the learners' job activities.
- Examining the extent to which training content has been applied to work practices.
- Identifying workplace factors which affect the extent of learning transfer.
- Gaining information about learners' feelings and attitudes.
- Providing learners with a vehicle to consider their own personal development goals and achievements.
- Comparing the congruence between organizational strategy and training provision.
- As a preparatory aid in the drafting of questionnaires.

Although interviewing has been described quite aptly as 'holding a conversation with a purpose' it is a technique which requires skill and preparation far beyond the levels required for a casual chat. It may be informal in tone but the interview is carefully structured to ensure that relevant information is collected from the interviewee and that the interviewer creates as little bias as possible in the replies. The quality of design and the skill of the interviewer make a significant difference to the quality of the data collected. A difficult problem to guard against is that poor interviewers are often not aware of the effect they have in obtaining superficial or distorted information. On the positive side, it is fair to say that people are not born interviewers: interviewing is a skill which can be acquired and practised like any other skill.

The stages through which the interview method proceeds are quite similar to those for questionnaire design:

1. Decide what kinds of information you need.
2. Draft the interview schedule.
3. Pilot the draft instrument and redraft as necessary.

4. Full-scale implementation.
5. Analyse the data.
6. Report the findings

WHAT KINDS OF INFORMATION DO YOU REQUIRE?

The recommendations made in Chapter 6 concerning questionnaires apply with equal force to interviews and will not be repeated here. However, the *number* of questions is a matter that does require extra emphasis. Interview questions are typically open-ended and respondents will more often ramble on at length than clam up tight. Twenty to forty minutes is a reasonable length to aim for, both in terms of the demand on the interviewee's time, and in terms of the evaluator's span of concentration. Serious listening is quite hard work. About 6 to 12 questions is a realistic maximum for that amount of time: more questions will mean that the interview becomes simply a verbal questionnaire and any depth in the responses will be lost.

DRAFTING THE INTERVIEW SCHEDULE

GENERAL REMARKS

The interview schedule consists of the list of questions which the evaluator will ask the interviewee. The point of having a schedule is to ensure *consistency*. The same questions will be asked every time, using the same words. However, the interview schedule should not be a strait-jacket for the interviewer. One of the particular strengths of interview as a method is the scope it creates for a two-way conversation with respondents. Thus, the main questions in the schedule will usually be supplemented both by prepared follow-up questions which probe for additional information and by off-the-cuff questions that arise naturally in the course of the interview. These questions which arise spontaneously may later be incorporated formally in the schedule, or they may prove to be specific to only one or two respondents.

SEQUENCING

The interview should start with an explanation of the purpose of the interview, the roles of the interviewer and interviewee, the sponsorship of the study, and the extent to which the interview will be confidential. This is also the point at which to reach agreement on the method of recording the interview, whether pencil and paper or tape.

Sequencing of questions should follow the same principles as for questionnaires: group together those questions that concern a similar theme; move from the general and the innocuous to the specific and the controversial; build the questions into an

order that makes psychological sense to interviewees; ask questions which help respondents to recall the pertinent facts before you ask questions about their feelings towards those facts.

TYPES OF QUESTION

There are two kinds of questions which should be avoided: leading questions and rhetorical questions. Both are more easily committed in the free-flow of an interview than in the more controlled context of a questionnaire.

Rhetorical questions

These are comparatively easy to avoid. They are statements, made in the form of a question, but which do not expect an answer. Typically, the person who poses the rhetorical question then goes on to provide her or his own answer, for example:

'Can we afford not to exhibit at Birmingham...?'

'Is this the face that launched a thousand quips?'

Leading questions

These are often similar in form to rhetorical questions but the intention is different. Leading questions are posed so as to get someone else to answer in the way preferred by the questioner, for example:

'Do you really think that we can justify that sort of expenditure on training?'

'Most people think that the right way to conduct training is like this. What is your view?'

'T-group training is often unpopular with participants, isn't it?'

There are two types of question commonly used in both questionnaire and interview – open and closed questions – and one type that is specific to interviewing: the reflective question.

Open questions

These are most commonly used in interviews and their use should be minimized in questionnaires. Open questions are valuable as a way to get an interviewee to talk ('Please tell me about the kinds of work you do') and as a way to discover the opinions, assessments and personal frame of reference of the respondent ('How did you feel about a training experience of that kind?'). A mixture of open and closed questions allows the interviewer to explore a complex subject by means of a series of questions that gradually narrow the focus of attention, without, however, leading the respondent into any preferred answer. For example:

Q8. 'Did you draw up any sort of plan – either on paper, or informally in your own mind – to put into practice things that you learned on the course?'

(closed question)

(if 'yes')

Q8.1. 'Please give me some specific examples of things from your plan that you have implemented in your job?'

(open question)

(if examples given)

Q8.2. 'Do you think these changes have made any difference to how effective you are in your job?'

(closed question)

(if 'yes')

Q8.2.1. 'What changes in your job effectiveness can you identify?' (If necessary, probe for specific instances.)

(open question)

(if 'no')

Q8.2.2. 'How do you feel about the fact that those changes you made seem to have made no difference to your job effectiveness?'

(open question)

It should be evident from these examples that the key feature of an open question is that there is no single answer to be expected from it, and answers will rarely be brief. It is because open questions can generate a range of responses that they are invaluable as a preliminary step towards design of closed multiple-choice type questions for use in questionnaires. An open question such as 'Please tell me about any preparatory briefing you received before the training event' may produce answers that build in to a multiple-choice item such as that which follows:

Q16. Indicate the nature of any pre-course briefing

- | | | |
|--|--------------------------|--------------|
| No briefing | <input type="checkbox"/> | (tick as |
| Written joining instructions | <input type="checkbox"/> | many |
| Discussion with training staff | <input type="checkbox"/> | boxes |
| Discussion with line manager | <input type="checkbox"/> | as are |
| Discussion with ex-participant of course | <input type="checkbox"/> | appropriate) |

Open questions carry a risk that the interviewee will use them as an opportunity to evade the question. Supplementary questions which probe for specific details will help to keep the interviewee on the subject in hand. They will also help to distinguish between those who are making general statements of good intent or wishful thinking and those who are talking about real actions and concrete events.

Closed questions

These questions are drafted in such a way that there can only be a short, very specific answer of a particular kind. For example:

'What is your job title?'

'How old are you?'

'Male/Female (*delete one*)'

'When did you attend the negotiating skills course?'

- 1981
- 1982 (*tick one*)
- 1983 *box*)
- 1984

A limitation of closed questions is that 'yes/no' type answers do not allow the respondent to qualify or expand the answer. Their value is, first, that very specific and unambiguous information is obtained and, second, that analysis of data is easier.

When drafting closed questions it is important to check that you are not creating either false dichotomies or 'leading' options. A false dichotomy arises where you might ask:

'When you identify training needs, do you mainly rely on job analysis or do you mainly respond to managers' requests for training?'

This question excludes valid alternative methods of needs identification, for example critical incident analysis. A closed leading question might ask, for example:

Q19. Please indicate how useful you found the session on 'selling to major accounts'	<input type="checkbox"/>
Very useful	<input type="checkbox"/>
Useful	<input type="checkbox"/>
Moderately useful	<input type="checkbox"/>

This example excludes the possible response 'Not at all useful' and leads the respondent towards the biased conclusion that the session *must* have been useful in some degree.

The comments on design of questions for use in questionnaires (pp. 82-9) apply with equal force to interviews and will not be repeated here. The advice on double-barrelled questions is worth re-emphasizing, however, because it is easier to create multiple questions in the free conversational flow of an interview than in written form ('Did you make any changes following the course, and what did your boss think about it - was it all worthwhile?').

Reflective questions

These form a useful part of the interviewer's repertoire. A reflective question consists of a brief re-statement of something that the interviewee has said, given with a questioning intonation in the voice. For example:

'You said that there were some problems with individual reactions to the trainers . . .?'

'It will sometimes be advisable to phrase the reflective statement tentatively so that the interviewee does not feel threatened.'

'I think you said there may have been some disagreement with your boss about your attendance on that particular course . . .?'

Reflective questions are useful in three ways. First, they are a relatively subtle way of probing into potentially controversial areas. Second, they allow interviewers to check whether or not they have correctly understood the respondent's statements. The reflective paraphrase gives the respondent the chance to correct any misunderstandings. Third, the reflective question demonstrates that the interviewer is paying attention to the respondent.

One final point on design of questions: it is a good rule to rely on Rudyard Kipling's 'honest serving men': What, Where, When, Who, and How¹. It is generally advisable to avoid 'why'. The difficulty with questions that begin 'why' is that very often the respondent will fall back on defensive explanations and rationalizations, rather than provide useful factual information about the circumstances under which something has occurred. Do not ask

'Why did you not make use of course learning back in your job?'

This will commonly produce answers that begin 'I have been too busy . . .' or 'It was all very well in theory but . . .' or the almost traditional response of 'My boss won't let me . . .'. Instead ask

'What factors in your own work situation have affected how you are making use of the course learning?'

THE INTERVIEW SCHEDULE: TWO EXAMPLES

Although there is no universally-applicable interview schedule, many questions will be common to different evaluation situations and can be easily adapted to match your own requirements from the examples given in Figures 7.1 and 7.2. However, you should always be prepared to draft your own questions so that you can gather data that will be relevant to the particular interests and criteria of worth of your audiences for evaluation.

PILOTING THE DRAFT SCHEDULE

DEVELOPING INTERVIEWING SKILLS

Interviewing is, of course, a practical skill. You will get better by doing it and by reflecting on your successes and the inevitable mistakes that arise. There are some general principles to guide you and two exercises are provided (Figures 7.3 and 7.4) for you to try out in a 'safe' environment.

The golden rule of interviewing is to *listen*, not talk. At least 95 per cent of the verbal

1. Please give me a brief outline of your job.
2. Have you taken part in any other training activities during the last three years? (*if yes*)
 - 2.1 Looking back over the various training activities you have taken part in, can you tell me what you take into consideration when you decide whether or not an event was good, bad, or whatever?
3. If you think back to before you attended course X, can you recall what you hoped to gain from it?
4. What, if anything, do you think you actually gained?
5. Can you describe any actual event – preferably recent – in which you acted differently because of what you learnt on the course? (*if not*)
 - 5.1 Does that mean there have been no changes in your job behaviour following the training? (*Probe: Do these answers appear to contradict answers given to Question 4?*) (*if yes*)
 - 5.2 What exactly happened? (*Probe: Who was involved; when; where; what was significant?*)
6. Do you see that course as having any impact on how effectively you do your job? (*if yes*)
 - 6.1 Can you please describe some specific examples where this has happened?
7. Did your manager do anything to prepare you for your attendance at the training event (*Probe for specifics*)
8. Did your manager do anything after the course to help you make use of the course content?
9. Are there any other aspects of your job where you feel training might be useful? (*Probe for details*)
10. What are your views on the methods the trainers employed during event X? (*Probe for details*)
11. Is there anything else about that course on which you would like to comment?

Figure 7.1 Specimen interview schedule questions: to ex-participants

1. Can I first confirm that you nominate members of your staff for various training events?
2. Can you please think about the range of training events for which you make nominations, and tell me what you take into consideration when you decide to nominate someone for a particular event?
3. Thinking specifically about course X, do you recall what you hoped – before the event – would be the benefits for the person you nominated?
4. Did you do anything before the course to prepare that person for it?
5. Did you do anything after the event to review what they had learned (*if yes*)
 - 5.1 Did you discuss how they might put the things they had learned into practice?
6. Do you feel that the time and money that course X incurred were a good investment or a waste of resources?
 - 6.1 Can you tell me what you took into account in making that judgement?
7. Is there anything else that you would like to comment on regarding course X?

Figure 7.2 Specimen interview schedule questions: to nominating managers

This exercise requires that you work with someone else – a colleague, friend, or partner. It is preferable to select someone with whom you have worked before and with whom you will not feel anxious.

The purpose of the exercise is to improve your listening skills by providing you with an opportunity to listen to views with which you will strongly disagree.

1. Name a subject about which you hold strong opinions. It should not be a technically specialized issue, requiring a high level of expertise, but should be a controversial matter of general knowledge or 'current affairs'.
2. Your partner should now prepare to talk for three or four minutes, arguing the case *against* your preferred opinion as strongly and convincingly as possible. As your partner delivers this statement you should listen, without comment, and make mental notes of all the significant points.
3. When your partner has finished, give a full and unbiased summary of what your partner has said. Ask your partner to tell you frankly how accurate you have been. Discuss your own feelings whilst listening to the point of view with which you strongly disagreed, and any difficulties these feelings created for you when you were required to provide an objective summary (and not a counter-argument).

Figure 7.3 Hearing what you don't want to hear

output should come from the interviewee. Interviewers should restrict their statements to the scheduled questions plus follow-ups and probes. Although the interview should start with a brief description of the purpose of the evaluation, this should never extend to offering comments on the anticipated outcomes of the research. Nor should the interviewer be drawn into debating views which the respondent advances. People are generally adept at telling you what you apparently want to hear. Do not give away clues.

Effective listening is a far from passive activity. It requires the interviewer not just to hear but to confirm the accuracy of his or her understanding, to clarify any ambiguities, and to explore – with a delicate touch – any apparently contradictory statements. Reflective summaries are invaluable for these purposes.

Effective listening also means hearing the bad as well as the good. To succeed as an interviewer you need to be willing to explore the respondent's point of view without letting your own biases and preconceptions act as filters. Figure 7.3 presents an Exercise, 'Hearing what you don't want to hear', which sets out a structure to practise listening to views with which you would usually disagree.

Tape-recording of interviews is often thought to be a controversial technique. The objections, in my experience, are far more likely to come from inexperienced interviewers than from interviewees. A handful of refusals in a decade of interviewing does suggest that taping is usually acceptable provided it is introduced in the right way. The initial contract of confidentiality must be talked through, before taping starts. It is usually easier to give credible guarantees of anonymity when the evaluator is recognized to be an independent outsider rather than a member of, say, the corporate personnel department. It is particularly important to state explicitly who has access to the recorded tape (preferably only the interviewer) and what will happen to the tape itself when the evaluation is completed. The norm should be that tapes are electronically wiped after data analysis; tapes should never be made available to the client if promises have been made to protect the identity of respondents.

The advantages of taping are as follows:

- Better attention and eye contact.
- The ability to capture all data, including nuances of tone and emphasis.
- Much greater freedom for the interviewer to follow up responses with supplementary questions – instead of busily writing notes, the interviewer can be an active participant in a two-way dialogue.

The main disadvantage of taping is cost – principally the time that is required to make even selective transcriptions of what has been said in the interview. When transcribing tapes allow yourself two to three times as long as the time occupied by the interview itself.

Listening goes hand-in-hand with using the power of silence. When you have asked a question, do not say another word. Wait. Silence *always* seems longer and more uncomfortable for the person asking the question than it does for the interviewee. Silence gives the interviewee time to think about the answer; it creates a pressure to give an answer. Resist the urge to restate the question or, worse still, to start to answer it yourself. Only if the interviewee explicitly asks for clarification should you restate or paraphrase the question. Finally, remember during silences to look attentive, maintain eye contact and create the impression that you are keen to receive the interviewee's reply.

There is a second consideration on the subject of silence: make it a rule not to interrupt the flow of the interviewee's replies. Interruptions can easily irritate the respondent. They are also likely to interfere with a line of thought which then gets lost beneath the new conversational direction. The only exceptions when interruption may be necessary are where the person has either misunderstood the question or is rambling in an unambiguously irrelevant fashion. More usually, simply make non-committal, vaguely encouraging noises of the 'Mmmm' kind.

One final observation on the importance of listening: the greatest value of the interview technique is its flexibility. You can probe and explore in response to what the interviewee says. Often the clues to valuable insights and opinions may be quite tentatively stated. There may be a passing allusion, a 'throwaway remark' which can be easily missed or disregarded. Careful listening – noting the tone as well as the words, the hints and undercurrents, even the things that do not get said – is essential if the more subtle kinds of interview data are to be gathered.

TESTING THE SCHEDULE

The interview schedule should be tried out on both subject matter experts and on a few people who are similar to your target interviewees. As the pilot interviews proceed, check what the respondents understand by the question and how they feel about answering it. Find out how respondents would have asked that question. Probe to discover whether your question has fully covered the subject. Delete and revise

1. Form a group with two other people to practise your interviewing skills. Use the schedule which you have drafted ready for pilot implementation. You will act as interviewer, one of your colleagues will be the interviewee and the other will act as observer. Set a time limit of 15–30 minutes (depending on the length of your draft schedule).
2. Conduct an interview in as realistic a manner as possible. The observer should make notes as the activity proceeds, using the checklist which follows.
3. When the interview is completed, invite both the interviewee and the observer to give you feedback on your performance. Discuss any problems arising.
4. Write down without argument any suggestions for how you can improve your interviewing technique. Resist the urge to play 'Yes, but...'. You are doing this in order to make yourself a highly professional evaluator. Welcome constructive criticism. 'The wise man always throws himself on the side of his assailants. It is more his interest than it is theirs to find his weak point' (Emerson).
5. Reflect later on the comments you have received. Not all of them will be accurate or helpful, but some almost certainly will be.
6. Repeat the exercise with a different interviewer and perhaps a different observer.

Observer's Checklist

<i>Use of questions</i>	<i>Frequency of use</i>	<i>Specific instances</i>
Open		
Closed		
Reflective		
Leading		
Other		
<i>Use of interview behaviour</i>	<i>Frequency of use</i>	<i>Specific instances</i>
Inaccurate listening		
Reflective summarizing		
Use of silence		
Interrupting the interviewee		
Clarification of ambiguity or contradictions		

Figure 7.4 Interview practice

questions as necessary. Remember that an interview schedule cannot include the number of questions that are possible with a written questionnaire.

Interviewers should also note their own reactions during the interview. What difficulties of access to interviewees were encountered? Was the reception unfriendly? Did certain questions provoke resistance or embarrassment? Were there any problems in maintaining a good working relationship with the interviewee? Were there any signs of irritation, impatience, or boredom in the interviewee? Was there a tendency to argue or talk too much on the part of the interviewer? Were there problems in recording data during the interview?

FULL-SCALE IMPLEMENTATION

ADMINISTRATIVE ASPECTS

Interviewing is best conducted on the interviewee's home ground. Prior appointment is essential. Travelling time and possibly overnight accommodation must be included in plans. An evaluation interview is normally held on a one-to-one basis: it should not be confused with a group discussion. Even where both manager and subordinate are interviewed concerning the latter's participation in a training event, they should be seen individually.

An introductory letter should be sent to potential interviewees, even before appointments are fixed, to inform them of the purpose of the evaluation study and to explain their own role within it. The letter should invite their participation rather than assume it. Few things are more futile than conducting an interview with an unwilling and hostile respondent. The introductory letter should also discuss the question of confidentiality of information.

A confirmatory letter may be sent when the interview has been agreed in principle and a date fixed. This letter, or a similar telephone communication, should mention the importance of being able to talk in private, rather than in an open office. It is imperative that the interviewee is cut off from his or her telephone for the duration of the meeting: few things are worse than the telephone for interrupting a line of thought or for interfering with the rapport established between the two parties.

At this initial stage the interviewee should be given a fair indication of the length of the meeting. This will be known quite accurately as a result of the pilot testing of the schedule. As a rule, you should stick to the maximum length you have quoted, unless the interviewee expressly indicates a willingness to continue. Respondents do vary in their talkativeness but you should over-run only rarely provided the piloting was done with representative subjects.

OPENING THE INTERVIEW

The first thirty seconds of any one-to-one encounter are critical in determining the direction in which it proceeds thereafter. Greet the interviewee by name, smile, intro-

duce yourself, and allow the interviewee time to invite you to sit down. It is a sensible precaution to match your own style of dress to that of the interviewee: people tend to like people who appear like themselves. In some situations age, sex, or ethnicity may present potential barriers and these occasions may call for interviewers with characteristics that match those of the respondents.

A few minor pleasantries break the ice and help to start the interviewee talking. You may then remind the interviewee of the purpose of the evaluation study and the agreement on confidentiality. A form of words that covers many evaluation interviews is:

'What you say will be used to help me compile a report to (the training manager) on (the supervisory skills programme). However, nothing you say will be used in a way that identifies the source of the information. Our discussion today is confidential in the sense that no one else will be able to find out what you have said.'

If you are using a tape-recorder, you may add (after discussing the use of the apparatus):

'Nobody but me will hear the tape and all the tapes are electronically wiped once I have transcribed the data I require.'

It is often very easy to move on the conversation from the statement of confidentiality to the idea of tape-recording it. There is far less resistance to use of recorders than you perhaps imagine and there are three very persuasive reasons to employ them. The first is that tape-recording allows interviewers to give all their attention to the respondent – maintaining eye contact and an attentive posture, instead of hiding behind a clipboard and scribbling away furiously. It seems likely that the sight of every word being painstakingly noted is rather more inhibiting than a discretely positioned, small, directional microphone linked to a miniature tape-recorder.

The second advantage of taping is that you capture *all* the data. With note-taking, you are constantly under pressure to keep up. Consequently, you tend to note the statements that fit into the preconceived framework of ideas with which you arrived at the interview. It is very easy to miss significant issues if they do not fit in with your expectations about the kinds of answer you will receive. A tape can be played back without this pressure and all ideas, even the unexpected, can be transcribed.

The third advantage of taping is that you can listen full-time to the interviewee and therefore can probe and clarify as the need arises. You can also be more sensitive to nuances of tone or content which hint at issues that need to be explored more fully.

THE MAIN BODY OF THE INTERVIEW

Work your way through the interview schedule in sequence of questions. It is not uncommon for respondents to answer a question that you have not yet reached, in the course of answering one you have asked. Where this happens, it is advisable to ask the later question anyway, when you come to that point in the schedule. You might

sometimes add an informal comment such as: 'I think we may have covered some of this ground already, but I'd like to ask the next question to be sure.'

The central skills of the interview are listening and maintaining a good relationship or 'rapport' with the interviewee. Rapport is very quickly dissipated if you start to argue about things the respondent has said, or if you pass judgements on them. Rapport is also lost if your attention is seen to be wandering or you look bored or irritated.

Rapport is strengthened by eye contact and by verbal support, both in the form of encouraging noises ('Uh-Huh', 'Yes ...', etc.) and in the form of short reflective summaries of what the interviewee has said. Rapport is strengthened in particular by your expressed acceptance at face value of what the interviewee says. Clear discrepancies in what is said at different points in the interview should be tactfully clarified. Catching-out untruths, or expressing disbelief in the interviewee's statements very rapidly undermines the working relationship and leads to defensive formality or termination of the interview.

However, it is also necessary to gather data that will illuminate the truthfulness (or lack of it) in the interviewee's answers. This can be done least threateningly by asking for concrete examples of situations that illustrate a particular claim to competence. This probing can be presented as a request for clarification or for additional detail. Be sensitive to three indicators of truthfulness in respondents' answers:

- Internal consistency – statements do not contradict each other.
- Balance – personally unfavourable information is provided, as well as that which puts the interviewee in a good light.
- Exaggeration – personal achievements are overstated or overvalued.

Last, remember that the guidelines on effective listening (pp. 103–04) apply to 'real' as well as pilot interviews.

When asking supplementary and probing questions, it sounds less interrogatory to preface your questions with 'May I ask ...' and similar phrases. Take care with off-the-cuff questions that you do not ask confusing, double-barrelled questions such as 'Did you find the lectures helpful or would you have preferred more role plays?' Be particularly careful that you do not inadvertently ask leading questions – when improvising probes, it is very easily done.

TERMINATING THE INTERVIEW

It is advisable to end with a broad 'catch-all' open question that gives the interviewee a chance to say anything that has not been covered previously. This question might be as broadly phrased as 'Is there anything else you want to say?' It may probe for topics missing from the schedule: 'Is there anything else that you think I should have asked you on this subject?'

The interviewee should be given an opportunity to ask any more questions about the purpose of the evaluation, confidentiality, and the like.

Lastly, interviewees should be thanked for their cooperation and the thought they have put into their answers. A modest exaggeration of the value of individual contributions may be justified for the goodwill it establishes towards the evaluation project.

DATA ANALYSIS AND REPORTING YOUR FINDINGS

These topics are considered in relation to the analysis of data from questionnaires and from interviews.

There are four kinds of conclusion that an evaluation study using questionnaire or interview may yield:

- Illustrative insights (typically arising from an individual's descriptive anecdote, or constituting a case study).
- Frequency statements (compiled from questionnaires or interviews, using a data matrix).
- Statements of co-variance (obtained by a cross-tabulation of different variables within the study).
- Statements of causal relationship (based on statistical, logical and factual evidence).

ILLUSTRATIVE INSIGHTS

These can be of great value to the training evaluator. The value of an idea lies in its fruitfulness – its potential utility. A good idea never needs to be 'representative' in any statistical sense.

It is when the evaluator is exploring trainee or manager perceptions, attitudes, feelings and beliefs about a particular learning event, or learning method, that this kind of insight is most likely to arise. It is characteristic of interview and of open-ended questioning.

Very little in the way of analysis is required. Statements or anecdotes may need to be depersonalized in order to respect confidentiality. Where more than one person has the same insight, it may be worth adding a frequency count to the substantive data, but as a rule illustrative insights stand or fall on their intrinsic quality, not their numerical quantity. Incidentally, there is research evidence that descriptive (case-study type) data is particularly influential with policy-makers. Its concrete reality has more persuasive impact than purely statistical information. Certainly, if this data has been collected in a systematic way, there is no reason to apologize for its non-numerical nature.

Please indicate the importance of these topics to success in your own job				
Topics	Importance			
	None	Some	A lot	No response
1. Handling face-to-face customer inquiries	HHH II ⑦	HHH HHH HHH II ⑯	HH HHH HHH HHH HHH HHH HHH III ⑳	III ④
2. Handling customer telephone inquiries	II ②	HH HHH HHH HH HHH HHH HH III ⑳	HH HHH HHH HHH III ⑳	II ②
3. Etc.				

Figure 7.5 Data matrix (assuming 66 respondents)

FREQUENCY STATEMENTS**Classifying data**

The first step in data analysis is to classify and count the responses to questions. With closed questions, the classification categories will be contained in the question, which in turn has been developed from your hypotheses about the training which you are investigating. Categories *must* have the following characteristics:

- They must be defined so that there is no overlap or ambiguity in the border between two categories.
- The set of categories must allow for all possible responses that can occur.

For example if you hypothesized a connection between age and openness to learning new techniques, one of your questions would ask for respondents' ages. A category set in an occupational setting might be:

Under 21	21-30	31-40	41-50	51-60	61 & over
<input type="checkbox"/>					

The basic tool for creating frequency statements is the data matrix. This combines *variables* with *units of analysis* so that a cell is formed for each combination of a variable with a unit. Into these cells are added the values of the responses you have collected. The data matrix can sometimes be simply created by using a master copy of the questionnaire item, with larger boxes so that you can score response frequencies (see Figure 7.5).

Ratings of course presentation	Respondent						
	(1 = low) (6 = high)	A	B	C	D	E	etc
1. Clarity of verbal presentation		3	4	4	3	4	
2. Questions answered adequately		4	5	4	3	5	
3. Discussion encouraged		2	3	4	2	5	
4. Opportunities for practice		5	6	5	5	6	
5. etc							

Figure 7.6 Ratings of course presentation

When you use a questionnaire item in this way, do not forget to add an extra column to account for people who do not answer that particular question.

The matrix can be designed so that the analysis provides data on individuals as well as on the evaluation variables and the units of analysis (or ratings) applied to those variables (Figure 7.6).

Averages

Frequency alone may be a significant piece of information: in Figure 7.5 it must be cause for further inquiry as to why there were seven respondents for whom the topic of face-to-face customer relations was perceived to be of no job importance. However, it is also useful to employ the basic data as a way of characterizing what is 'typical' about the group of respondents. There are two measures – the average and the central tendency (the clustering around a midpoint).

In Figure 7.5, there are no grounds for assuming that the three scale items (None/Some/A lot) represent *equal* intervals along a scale. However, they do form an *ordered* measure from lesser to greater quantity. In this instance, the appropriate measure of the average is the *median*. When all scores are placed in order of magnitude, the median is the midpoint position. In Figure 7.5 the midpoint of 66 scores is 33. Adjusting for non-response cases, the median for Topic 1 will be the 31st case ($\frac{66 - 4}{2}$). Adding the totals in each cell, starting from the left (None) box, the 31st median case is the 7th case in the 'A lot' cell.

In Figure 7.6, the scores may be assumed to be based on an equal interval scale. It is therefore appropriate to use the *arithmetic mean* as the average. The data matrix would of course extend far beyond Respondent E and would include the whole of the sample (sometimes the whole of the population) included in the evaluation. In the simplified illustration here, the arithmetic mean for the first rated item (Clarity of verbal presentation) is 3.6. The mean is arrived at by adding together all the individual

To what extent are you involved in identification of training needs?		
a) Not involved	(6)	
b) Conduct detailed analysis	(22)	
c) Rely on managers' identification of needs	(10)	
d) Assist senior trainer to conduct analysis	(16)	
e) Other	/// (3)	

Figure 7.7 Analysis of nominal scale to illustrate the mode

ratings ($3 + 4 + 4 + 3 + 4$) and then dividing the total (18) by the number of cases (5 respondents).

The third measure of 'average' is the *mode*. This is the score which occurs most frequently. The mode is used for what are called nominal scales, where the various items do not form any sort of order of magnitude.

In Figure 7.7, the modal response is b) – 'Conduct detailed analysis'.

Dispersion

It is useful to know how widely individual cases vary from the average of a group. There are three measures of this:

1. *Range* is the simplest and states the extreme values of the group scores. In Figure 7.6, the range varies between topics – only 3–4 on Topic 1, and 2–5 on Topic 3, indicating a greater variety of respondent opinions.
2. To establish *quartile deviation* response values are placed in order of magnitude along their scale. Positions known as quartiles fall one-quarter and three-quarters of the way along the distribution. The quartiles are located half-way between the median value and the upper and lower extreme values.

Quartile deviation is calculated thus:

$$QD = \frac{Q_3 - Q_1}{2} \text{ where } Q_3 \text{ is the upper quartile and } Q_1 \text{ the lower.}$$

The quartile deviation is expressed in terms of the value units which form the ordered distribution. To compare QDs you should calculate the quartile coefficient of dispersion:

$$QCD = \frac{(Q_3 - Q_1)}{2} \div \frac{(Q_3 + Q_1)}{2}$$

3. *Standard deviation* shows the average distance of individual cases from the group

mean. Thus SD presupposes an equal interval scale is being used. It is calculated by the formula:

$$SD = \sqrt{\frac{\text{Sum of squares of individual deviations from arithmetic mean}}{\text{Number of items}}}$$

For example

Scores	Deviations from AM	Squares of deviations
4	-1	1
7	+2	4
3	-2	4
5	0	0
8	+3	9
2	-3	9
6	+1	1
Total	35	28

$$AM = 35/7 = 5$$

$$\text{No. of scores} = 7$$

$$SD = \sqrt{\frac{28}{7}} = 2.0$$

Comparing frequencies

A common requirement for evaluators is to compare the responses obtained from one group with those obtained from another. The comparison of scores (e.g. from written tests) is discussed in Appendix II, p. 246. The comparison of response frequencies in the answers to questions is useful as an indication of consistency or of change occurring over time, as between one course group and another. It is only in extreme cases that visual inspection of two sets of frequencies will show clearly that there is a difference between them (or not), and that inspection still would not tell you *how* alike or different the frequencies were.

The technique for comparing frequencies is the chi-square test (pronounced Ky to rhyme with 'eye', and written as χ^2). χ^2 tests whether the frequencies in a distribution differ significantly from some other set of frequencies. It is calculated by the formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O = observed frequency and E = a frequency which could be expected, on the basis of some assumed hypothesis about the subject matter. The calculation proceeds as follows:

Step 1: responses to a question are summarized for two or more groups, as below.

Q16. How frequently do you make use of the technical reference materials provided on the course?					
At least once a day (A)	At least once a week (B)	At least once a month (C)	At least once a year (D)	Never (E)	
Group I	1	3	10	4	1
Group II	0	5	7	1	0

Step 2: The row, column and overall totals are calculated, thus:

	A	B	C	D	E	Row totals
Group I	1	3	10	4	1	19
Group II	0	5	7	1	0	13
Column totals	1	8	17	5	1	32

The sum of the row totals equals the sum of the column totals and is referred to as the overall total; 32 in this example.

Step 3: For each cell in the matrix, an 'expected value' is calculated by means of the formula:

$$\frac{\text{Row Total} \times \text{Column Total}}{\text{Overall Total}}$$

For example, for Group I, Column C, the expected value becomes:

$$\frac{19 \times 17}{32} = 10.09 \text{ (rounded to } 10.1\text{)}$$

Applied to the whole matrix, the expected values are:

	A	B	C	D	E	Row totals	
Group I	Observed	1	3	10	4	1	19
	Expected	0.6	4.8	10.1	3	0.6	
Group II	Observed	0	5	7	1	0	13
	Expected	0.4	3.3	6.9	2	0.4	
Column totals		1	8	17	5	1	32

Step 4: Subtract each expected value from its corresponding observed value within its group. Square the result. Then divide that figure by the expected value and total the column.

Observed values	Expected values	O-E	(O-E) ²	$\frac{(O-E)^2}{E}$
1	0.6	0.4	1.6	2.67
3	4.8	1.8	3.24	0.68
10	10.1	0.1	0.1	0.009
4	3	1	1	0.34
1	0.6	0.4	1.6	2.67
0	0	0	0	0
0	0.4	0.6	3.6	9
5	3.3	1.7	2.89	0.88
7	6.9	0.1	0.1	0.02
1	2	1	1	0.5
0	0.4	0.4	1.6	4
0	0	0	0	0
$\chi^2 =$				20.769

Step 5: The χ^2 value of 20.77 gives a measure of the difference between the two groups. This difference may, however, be due to chance and it is important to assess whether or not this is a probable factor in the situation. This is achieved by looking up the critical value for χ^2 in a table of critical values tabulated against 'degrees of freedom'. You obtain the df value thus:

$$df = (\text{no. of rows minus 1}) (\text{no. of columns minus 1})$$

In the example above, $df = (2-1)(5-1) = 4$. Tables show a critical value for χ^2 of 9.49 where $df = 4$, subject to the assumption that there is no worse than a 1 in 20 probability of the results being due to chance. It depends on how important it is for your confidence in the results as to whether you select a 5 per cent level of confidence, a

1 per cent, or even an 0.1 per cent level (i.e. a value of χ^2 that would be exceeded by chance only once in a thousand similar analyses). There are published tables for χ^2 values which show critical values for a range of levels. At the 5 per cent level for df 1 to df 20 the values are as follows:

<i>df</i>	<i>Critical value for χ^2</i>
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92
10	18.31
11	19.68
12	21.03
13	22.36
14	23.68
15	25.00
16	26.30
17	27.59
18	28.87
19	30.14
20	31.41

In the worked example, χ^2 was calculated as 20.77. This puts it well above the critical value of 9.49 and enables you to state that the two frequencies are different, with no more than a 1 in 20 probability that the differences are due to chance.

STATEMENTS OF CO-VARIATION

Statistical correlation arises where two or more groups of values vary together, either directly or inversely. The coefficient of correlation gives a measure of this co-variance. The formula can be summarized as:

The mean product of the deviations from the mean

(The standard deviation of the 1st group)(The SD of the 2nd)

or as:

$$r = \frac{n \sum XY - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] [n \sum Y^2 - (\sum Y)^2]}}$$

where r is the Pearson correlation coefficient

n is the number of cases

X is the first set of values attaching to the cases

Y is the second set of values attaching to the cases

If the correlation is perfect and direct, the coefficient will be +1. A perfect inverse correlation yields -1. A coefficient of 0 suggests a wholly random connection between the two sets of values. Most coefficients will fall in the mid-range between -1 and +1. If a coefficient has a value of 0.30 or less then the correlation is weak. A level of 0.50 or higher should be required before making statements of correlation. For a worked example of the calculation of the coefficient of correlation see Appendix II, p. 254).

A common use of correlation is to show that some aspect of training, for example test scores, varies directly with some characteristic of the learners, for example level of previous education. However, it is a misconception of the nature of correlation then to state that the level of previous education determines or causes the level of test scores. Correlations permit a prediction but do not prove a causal connection.

STATEMENTS OF CAUSALITY

If you wish to say that X causes Y (e.g. that exposure to higher education leads to the use of impenetrable jargon), you have to satisfy three conditions:

1. That X and Y vary together (values of X and Y can be correlated); and the way in which they vary together can be specified from your hypothesis about that relationship.
2. That Y (the effect) occurred later in time than X (the cause).
3. That there were not other factors in the situation, besides X, that caused Y.

Requirement 1 has been discussed in the previous sections. Requirement 2 is a matter of observation. Requirement 3 is tackled by applying existing knowledge and logic to the hypothesized relationship of causality ('does "X causes Y" make sense in terms of what we know of the world?'). You must actively search for other explanations and in particular for spurious relationships. These arise where the apparent connection between X and Y is in fact due to both co-varying with a third variable W:



ANALYSING RESPONSES TO OPEN-ENDED QUESTIONS

Step 1

Review the raw data. If it was collected in an interview, read through your notes or (preferably) play-back the tape of what was said. If the data consists of answers to open questions on a questionnaire, read through the answers.

Step 2

Look for key words or phrases which accurately summarize what the respondent is saying. Be very careful if you have to provide your own summarizing word for a lengthy description given by the respondent. The risk is that you will end up analysing your own opinions rather than those of the respondent.

Step 3

Key words should directly relate to the question and hence to your purpose in conducting the evaluation. As you proceed through the data, start to group the words you have highlighted into categories (which must, of course, be both internally coherent and mutually exclusive).

Step 4

With clear response categories established you can proceed with data analysis as described in the preceding sections.

REPORTING YOUR FINDINGS

Reports on evaluation of training are, by the nature of the subject, rarely wholly favourable or wholly unfavourable. Reality often turns out to be a bit fuzzy around the edges. The practical effect of this is that there can be some tension between decision-makers' preferences for clear-cut findings, and an evaluation report which is accurate but qualified by some reservations or by a proportion of mixed conclusions. It is therefore most important that the evaluation has been conducted with a clear grasp of the specific criteria of worth which are significant to the decision-makers.

Begin an evaluation project by identifying your audience and their criteria for judging the training activity which you are about to evaluate. You are then unlikely to produce a mismatch between expectations and actual results. This mismatch is common in evaluation projects where the strategy has not been properly developed. It quickly undermines the credibility of evaluation as a worthwhile activity and makes adequate resourcing unlikely.

What this means is that you should start your evaluation report by stating for whom it has been compiled and on what criteria of worth judgements have been made.

The main body of your report will contain the findings of the evaluation study. Some information will be reported as summary statements, as in Figure 7.8. According to your purpose, you might then go on to show correlations between pre-course briefing (or its absence) and amount of learning achieved, or participants' reactions to course methods and content.

Some information may be presented most conveniently in tabular form or using visual summaries such as graphs, pie-charts or histograms.

When you quote summary data, always explain what it signifies in non-technical language. Aim for brevity, clarity and plain English. Wittgenstein said it all with: 'Anything that can be said, can be said clearly.'

Interviews were conducted with 61 level D managers who had nominated staff for the 'Contractor liaison and negotiation' programme. This represents 83 per cent of managers who make nominations for this programme.

When asked about how they prepared staff for participation in the programmes, responses were as follows:

Discussion at annual appraisal	8	(13.1%)
Discussion shortly before the programme	9	(14.8%)
No discussion and/or passed formal briefing papers to participant	30	(49.2%)
Arranged briefing by training department or by other ex-participant of the programme	13	(21.3%)
No answer given	1	(1.6%)
	61	100%

Figure 7.8 Evaluation report summary statement

A SUGGESTED FRAMEWORK FOR THE EVALUATION REPORT

1. Contents, plus introductory remarks
2. Statement of intended audience
3. Statement of evaluation criteria employed
4. Summary of conclusions and recommendations
5. Main text:
 - a) Summary of each finding
 - b) Interpretation of each finding
 - c) Action implications of finding
6. Concluding remarks
7. Appendices:
 - a) Description of the methods employed to gather data, the sample used, etc.
 - b) Raw data from which the summaries have been drawn
 - c) Extended case studies (if applicable).

NOTE

1. Kipling, R. (1940, reprinted 1973), 'The Elephant's Child', *Rudyard Kipling's Verse Definitive Edition*, Hodder & Stoughton.