EVALUATING IN-SERVICE PROGRAMMES FOR LANGUAGE TEACHERS

Ian McGrath

PhD

University of Edinburgh

1997

DECLARATION

I declare that this thesis has been composed by myself and that the work involved is entirely my own.

Ian McGrath
Nottingham, July 6, 1997

ACKNOWLEDGEMENTS

I was prompted to set out on this journey by my wife, Natasha. Almost five years on, I am still grateful to her.

My thanks are also due to Eric Glendinning, Director of the Institute for Applied Language Studies at the University of Edinburgh, whose enlightenment and flexibility with regard to study leave enabled me to pick up a small head of steam and keep going; to my first supervisor, Alan Davies, for his light but sure touch in the early years, and to Hugh Trappes-Lomax, who agreed to take on the role of 'critical friend' in the later stages and fulfilled that task with great conscientiousness and the utmost tact. Without Hugh's guidance, the journey might have been shorter, but I would have made more wrong turnings and seen a good deal less.

I also owe a considerable debt of gratitude to the many others who took the time and trouble to answer my questions or helped in other ways: all those who responded to the survey reported in Chapter 4, and especially the five anonymous informants who granted me interviews; the colleagues at IALS who cooperated in the case study reported in Chapter 5, and those named and unnamed whose evaluations of TLE appear in Chapter 6; Peter Wheeldon, whose wonderful openness made Chapter 7 possible, and Colin Smith and the participants in course 93T130 for their tolerance; the members of the 'Sprenger group', for their contributions to Chapter 8; and Charles Alderson, for a detailed and informative letter. Philip Goertzen, and John McEwan provided the technical back-up without which any explorer can feel very lonely.

ABSTRACT

The thesis is concerned with the evaluation of in-service programmes for language teachers. The main focus is on UK-based face-to-face provision for teachers of English to speakers of other languages, although some reference is made to programmes for teachers of languages other than English.

A number of questions are posed:

- What procedures are currently used by UK institutions to evaluate in-service programmes for language teachers ?
- How satisfactory is the approach to evaluation in use in UK institutions in the eyes of key respondents and judged by such criteria as validity and reliability ?
- What procedures other than those in common use might contribute usefully to programme evaluation ?

These questions are explored by means of survey techniques and case studies. On the basis of the resulting evidence, it is argued that there is a need for review of existing evaluation practices and for the dissemination of good practice.

The thesis is structured as follows:

Chapter 1 defines key terms, presents the case for programme evaluation and offers a rationale for the focus adopted. Chapters 2 and 3 deal respectively with theoretical issues and evaluation method and provide a context for the empirical work described in subsequent chapters. Chapter 4 reports the findings of a survey into current evaluation practices in UK institutions, and Chapters 5-8 comprise case studies in which the potential contribution of specific programme evaluation techniques is examined. Chapter 9 makes recommendations for future practice.

LIST OF ABBREVIATIONS

ESOL	English to speakers of other languages
INSET	in-service education and training
L2	second or foreign language
ML	modern languages
NS	native speaker
NNS	non-native speaker
PRESET	pre-service education and training
RAF	rapid assessment form
TESOL	teaching English to speakers of other languages
TL	target language
TML	teaching modern languages

LIST OF FIGURES AND TABLES

Figures

1.1	Focussing down	8
2.1	A decision hierarchy for evaluation	28
2.2	Contribution of summative and formative evaluation to evaluation purpose	33
2.3	Prioritising evaluation	43
2.4	An example of 'post-it' feedback on tutorials	56
2.5	The reliability and validity of interview data	68
3.1	The organisation and evaluative/pedagogic objectives of the study	106
7.1	Attendance figures, by subgroup	253
8.1	Action plans, page 1: briefing notes	267
8.2	Action plans, page 2	268
8.3	Revised lay-out for action plan	281
8.4	The action plan as stimulus to action and self-evaluation	284

Tables

1.1	Types of programme for language teachers	9
2.1	Methods for evaluating various objective types	64
3.1	Features of questionnaires	90
3.2	Features of interviews	96
3.3	Features of participant journals	108
3.4	Features of participant plans	115
3.5	Features of observation	119
3.6	Features of document analysis	126
3.7	Method use in L2 INSET evaluation studies	128
4.1	Breakdown of questionnaire returns by institution type	134
4.2	Completed returns, broken down by course type	138
4.3	In-course evaluation procedures used, showing rank order by frequency	141
4.4	End-of-course evaluation procedures used, showing rank order by frequency	144
4.5	Summary of post-course evaluation procedures used	144
4.6	Average number of measures used on different programme types	150
4.7	Tally of short responses to Q.5 on questionnaire	164
5.1	Pre-seminar questionnaire: extent of previous use of OHP and reasons for non-use	192
5.2	Pre-seminar questionnaire: purpose of OHP use	192
5.3	Pre-seminar questionnaire: attitudes to OHP use	193
5.4	Previous OHP use	194
5.5	Frequency of OHP use	194
5.6	Purpose of use: aggregated results	195
5.7	Nature of attitude change	197
6.1	Extract from Rapid Assessment Form	206
6.2	Extract from summary of RAF ratings for Methodology sessions: comparison across courses	208
7.1	Responses to statements by scale point and subgroup	246
8.1	Content areas featured in course and figuring in action plans	271
8.2	Action plan objectives related less directly to course content	271
8.3	Number of participants taking action on specific objectives related to course content	273
8.4	Number of participants taking action on objectives less directly related to course content	274
8.5	Individual implementation of action plans	276
9.1	The case studies: methods used and data collected	294

CONTENTS

Acknowledgements i
Abstract ii
Abbreviations iii
List of Figures and Tables iv

1. EVALUATING INSET PROGRAMMES FOR SECOND LANGUAGE TEACHERS: A RATIONALE 1

1.1 Introduction 1
1.2 The nature of programme evaluation 2
1.2.1 Programmes, projects and courses 2
1.2.2 Evaluation 3
1.2.3 Programmes and their evaluation 4
1.3 The need for programme evaluation 5
1.4 The specificity of INSET for language teachers 7
1.4.1 Introduction 7
1.4 2 INSET evaluation vs PRESET evaluation 10
1.4.3 INSET programmes for language teachers 11
1.4.4 Summary 12
1.5 INSET evaluation in the UK: a brief history 12
1.6 Evaluation in second language education 16
1.6.1 Introduction 16
1.6.2 The evaluation of programmes for language learners 16
1.6.3 The evaluation of programmes for language teachers 21
1.7 Research questions, method and structure of the thesis 22
1.8 Summary 25

2. PLANNING AND IMPLEMENTING EVALUATION 26

2.1 Introduction 26
2.2 A framework for evaluation 26
2.2.1 Key questions 26
2.2.2 A decision hierarchy 27
2.3 Evaluation purposes 30
2.3.1 Evaluation for accountability and development 30
2.3.2 Summative and formative evaluation 32
2.4 Evaluation criteria and foci 34
2.4.1 Introduction 34
2.4.2 The quest for suitable criteria 35
2.4.3 Criteria in project evaluation 37
2.4.4 Foci of evaluation 39
2.4.4.1 Levels of foci 40
2.4.4.2 Deciding on foci 42
2.5 The identity and role of the evaluator 44
2.5.1 Introduction 44
2.5.2 Participation 45
2.5.3 Insider or outsider evaluation 46
2.5.4 Subject specialist or generalist 48
2.5.5 The role of the external evaluator 49

2.6 The timing of evaluation 51
2.6.1 Evaluation in planning for INSET 51
2.6.2 Pre-course evaluation 53
2.6.3 Formal in-course monitoring 55
2.6.4 End-of-course evaluation 58
2.5.5 Post-course evaluation 59
2.6.6 Informal evaluation 62
2.7 Choosing a method 63
2.7.1 Introduction 63
2.7.2 A framework for the comparison of evaluation methods 65
2.7.2.1 Logistical factors 65
2.7.2.2 Methodological factors 67
2.7.2.3 Utility 70
2.7.2.4 Conflict between principles 71
2.7.3 Constraints 72
2.7.4 Politics, ethics and communication 73
2.7.4.1 Introduction 73
2.7.4.2 Politics 74
2.7.4.3 Ethics 75
2.7.4.4 Communication 76
2.8 Reporting 78
2.8.1 Style and content 78
2.8.2 The role of the evaluator-reporter 80
2.8.3 Circulation and participation 81

3. DATA COLLECTION METHODS AND EVALUATION 83
3.1 Introduction 83
3.1.1 Categorising data collection instruments 83
3.1.2 The best method 85
3.1.1 Combining methods 85
3.1.4 Triangulation 87
3.1.5 Evaluating programme evaluation methods 88
3.2 Questionnaires 88
3.2.1 Introduction 88
3.2.2 Questionnaire administration 89
3.2.3 Categorising questionnaires 89
3.2.4 Advantages and disadvantages of questionnaires 90
3.2.5 The lessons of experience 93
3.2.6 Conclusion 95
3.3 Interviews 95
3.3.1 Introduction 95
3.3.2 Advantages and disadvantages of interviews 96
3.3.3 Administration 99
3.3.4 Categorising interviews 101
3.3.5 Structure *vs* control 102
3.3.6 Conclusion 103
3.4 Participant journals 103
3.4.1 Introduction 103
3.4.2 Journals and programme evaluation 104
3.4.3 Advantages and disadvantages of participant journals 108
3.4.4 Conclusion 110

3.5 Participant plans 111
3.5.1 Letters 111
3.5.1.1 Letter to the trainer 112
3.5.1.2 Letter to the boss 112
3.5.1.3 Letter to myself 113
3.5.2 Action plans 114
3.5.3 Advantages and disadvantages of participant plans 115
3.6 Observation 116
3.6.1 Introduction 116
3.6.2 Observation as a means of evaluating implementation 117
3.6.3 Participant observation 118
3.6.4 Advantages and disadvantages of observation 119
3.6.5 Conclusion 120
3.7 Document analysis 120
3.7.1 Categories of document 122
3.7.2 Purposes of document analysis 123
3.7.3 Participant assessment 124
3.7.4 Advantages and disadvantages of document analysis 126
3.7.5 Conclusion 126
3.8 Conclusion 127

4. SURVEY 128
4.1 Introduction 128
4.2 Objectives 128
4.3 Selection of sample 129
4.4 The questionnaire 131
4.4.1 Introduction 131
4.4.2 The letter 131
4.4 3 The courses checklist 132
4.4 4 Design of the questionnaire 132
4.4.5 Return rate 133
4.4.6 Non-returns and null returns 134
4.4.7 Sample bias 135
4.4.8 Reliability 135
4.5 The interviews 136
4.5.1 Planning for the interviews 136
4.5.2 Conducting the interviews 136
4.5.3 Follow-up to the interviews 137
4.5.4 Analysing the data 137
4.6 Findings and discussion 137
4.6.1 Introduction 137
4.6.2 Classifying the data 138
4.6.3 Evaluation stages and procedures 139
4.6.3.1 Collection of baseline data 139
4.6.3.2 In-course monitoring 141
4.6.3.3 End-of-course evaluation 144
4.6.3.4 Post-course evaluation 144
4.6.3.5 Summary 145
4.6.4 Programme types and evaluation procedures 146
4.6.4.1 Degree-level programmes 146
4.6.4.2 Other award bearing programmes 148

4.6.4.3 Non-certificated closed-group programmes 149
4.6.4.4 Non-certificated open-entry programmes 150
4.6.4.5 Summary 152
4.6.5 Less commonly used procedures 152
4.6.5.1 Diaries 153
4.6.5.2 Post-course questionnaires 155
4.6.5.3 Feedback from people outside the course team 157
4.6.6 Factors affecting evaluation 159
4.6.6.1 Programme-type factors 159
4.6.6.2 Programme length/time available for evaluation 160
4.6.6.3 Type of institution 163
4.6.6.4 Institutional culture 163
4.6.7 Attitudes to evaluation 164
4.6.7.1 Responses to the questionnaire 164
4.6.7.2 The interviews 166
4.6.7.3 Interpreting the responses 168
4.6.8 Further issues 171
4.6.8.1 Participant opinions as evaluation data 171
4.6.8.2 Intended effects and criteria for assessing these 172
4.6.8.3 Use made of evaluation data 176
4.6.8.4 Modifications to evaluation procedures over time 177
4.7 Evaluation of survey instruments and procedures 178
4.7.1 Survey design and administration 178
4.7.2 Evaluation of the questionnaire 179
4.7.3 Evaluation of the interviews 182
4.8 Conclusions 183
4.8.1 General conclusions 183
4.8.2 Towards appropriate evaluation 184
4.8.3 Course-sensitive evaluation 186

5. CASE STUDY 1: A SEMINAR FOR ML TEACHERS 188
5.1 Introduction 188
5.2 The Study 189
5.2.1 Introduction 189
5.2.2 Evaluation questions 189
5.2.3 Stages in the study 189
5.2.4 Seminar participants 190
5.2.5 Seminar content 191
5.2.6 Results and discussion
5.2.6.1 Pre-seminar questionnaire 191
5.2.6.2 Post-seminar interviews 194
5.2.7 Conclusions 197
5.3 Implications of the case study 198

6. CASE STUDY 2: A REFRESHER COURSE FOR ESOL TEACHERS 201
6.1 Introduction 202
6.2 The study 202
6.2.1 Introduction: TLE and its origins 202
6.2.2 Overview of the case study 204
6.2.3 Evaluation procedures: 1982-1992 204
6.2.3.1 Individual tutorials 205

6.2.3.2 Rapid Assessment Forms 206
6.2.3.3 Tutor meetings 210
6.2.3.4 Study diary 210
6.2.3.5 First-week letter 211
6.2.3.6 Trouble-shooting sessions in Weeks 2 and 3 212
6.2.3.7 Participant self-evaluation 213
6.2.3.8 End-of-course questionnaire 214
6.2.3.9 Whole-group discussion 216
6.2.3.10 Course Director's report 216
6.2.4 Evaluating the evaluation procedures 217
6.2.4.1 Logistical factors 218
6.2.4.2 Methodological factors 219
6.2.4.3 Utility 220
6.2.4.4 Conclusions 222
6.3 Reviewing evaluation procedures 223
6.3.1 Introduction 223
6.3.2 The experiment: summer 1993 224
6.3.2.1 Background to the experiment 224
6.3.2.2 Results 224
6.3.2.3 Discussion 226
6.3.2.4 Conclusions 227
6.4 Implications of the case study 227
6.4.1 Answers to the research questions 227
6.4.2 The need for review 228
6.4.3 The need for documentation and standardisation 229

7. **CASE STUDY 3: A COURSE FOR ML PROBATIONERS AND RETURNERS** **230**
7.1 Introduction 230
7.2 The study 231
7.2.1 Introduction 231
7.2.2 Evaluation questions 232
7.3 Brief description of the course 232
7.3.1 Duration and intensity of course, location, participants, attendance 232
7.3.2 Tutors 233
7.3.3 Course aims 233
7.3.4 Course content 234
7.3.5 Facilities and resources 234
7.4 Evaluation procedures 234
7.4.1 Rationale for evaluation procedures 234
7.4.2 My role 235
7.4.3 Getting to know people 236
7.5 Journals 236
7.5.1 Introduction 236
7.5.2 Journal format 236
7.5.3 Presentation of the journal as an evaluation measure 237
7.5.4 Other factors affecting the study 238
7.5.5 Use made of the journals 238
7.5.6 How the journals were completed 239
7.5.7 Content analysis 240
7.5.7.1 Analysis strategy 240

7.5.7.2 Evidence of follow up 240
7.5.7.3 Evaluative comments 241
7.6 The questionnaire 243
7.6.1 Design of the questionnaire 243
7.6.2 Administration of the questionnaire 244
7.6.3 Plenary discussion 244
7.6.4 Results of the questionnaire: scale questions 244
7.6.5 Individual comments 250
7.7 Plenary discussion 251
7.8 Criteria for evaluation 252
7.8.1 Participant satisfaction 252
7.8.2 Fulfilment of programme aims 253
7.8.3 Other effects 255
7.9 Recommendations and implementation 256
7.9.1 Recommendations 256
7.9.2 Implementation of recommendations 257
7.10 Implications of the case study 258
7.10.1 Introduction 258
7.10.2 The journals 258
7.10.3 The questionnaire 259
7.10.4 Plenary discussion 260
7.10.5 Informal comments 260
7.10.6 The use of a non-participant observer 260
7.10.7 Triangulation 261

8. CASE STUDY 4: A CLOSED GROUP COURSE FOR ESOL TEACHERS AND TRAINERS 263
8.1 Introduction 263
8.2 The study 265
8.2.1 Course participants 265
8.2.2 Background to the course 265
8.2.3 Course content 265
8.2.4 Evaluation questions 265
8.2.5 Stages in the study 266
8.2.5.1 Writing the action plans 266
8.2 5.2 Follow-up 268
8.2.6 Data analysis 269
8.2.7 Results 270
8.2.7.1 Data analysis stage 1: the action plans 270
8.2.7.2 Data analysis stage 2: comparison of action plans and participant reports on action taken on these 272
8.2.7.3 Data analysis stage 3: explanations and other comments 275
8.2.7.4 The action plan 277
8.2.8 Conclusions 278
8.3 Implications of the case study 279
8.3.1 The action plan as programme evaluation instrument 279
8.3.2 Issues 280
8.3.2.1 Negotiation of action plans 280
8.3.2.2 Timing of follow up 281
8.3.2.3 Reliability of self-reports 282
8.3.2.4 Gathering data on cause and effect 283

7.5.7.2 Evidence of follow up 240
7.5.7.3 Evaluative comments 241
7.6 The questionnaire 243
7.6.1 Design of the questionnaire 243
7.6.2 Administration of the questionnaire 244
7.6.3 Plenary discussion 244
7.6.4 Results of the questionnaire: scale questions 244
7.6.5 Individual comments 250
7.7 Plenary discussion 251
7.8 Criteria for evaluation 252
7.8.1 Participant satisfaction 252
7.8.2 Fulfilment of programme aims 253
7.8.3 Other effects 255
7.9 Recommendations and implementation 256
7.9.1 Recommendations 256
7.9.2 Implementation of recommendations 257
7.10 Implications of the case study 258
7.10.1 Introduction 258
7.10.2 The journals 258
7.10.3 The questionnaire 259
7.10.4 Plenary discussion 260
7.10.5 Informal comments 260
7.10.6 The use of a non-participant observer 260
7.10.7 Triangulation 261

8. CASE STUDY 4: A CLOSED GROUP COURSE FOR ESOL TEACHERS AND TRAINERS 263
8.1 Introduction 263
8.2 The study 265
8.2.1 Course participants 265
8.2.2 Background to the course 265
8.2.3 Course content 265
8.2.4 Evaluation questions 265
8.2.5 Stages in the study 266
8.2.5.1 Writing the action plans 266
8.2 5.2 Follow-up 268
8.2.6 Data analysis 269
8.2.7 Results 270
8.2.7.1 Data analysis stage 1: the action plans 270
8.2.7.2 Data analysis stage 2: comparison of action plans and participant reports on action taken on these 272
8.2.7.3 Data analysis stage 3: explanations and other comments 275
8.2.7.4 The action plan 277
8.2.8 Conclusions 278
8.3 Implications of the case study 279
8.3.1 The action plan as programme evaluation instrument 279
8.3.2 Issues 280
8.3.2.1 Negotiation of action plans 280
8.3.2.2 Timing of follow up 281
8.3.2.3 Reliability of self-reports 282
8.3.2.4 Gathering data on cause and effect 283

8.2.7.4 The action plan 277
8.2.8 Conclusions 278
8.3 Implications of the case study 279
8.3.1 The action plan as programme evaluation instrument 279
8.3.2 Issues 280
8.3.2.1 Negotiation of action plans 280
8.3.2.2 Timing of follow up 281
8.3.2.3 Reliability of self-reports 282
8.3.2.4 Gathering data on cause and effect 283
8.3.2.5 The managerial dimension 283
8.3.3 The action plan as pedagogic device 284
8.3.4 Conclusion 284

9. SUMMARY AND CONCLUSIONS 286
9.1 Introduction 286
9.2 Case study research 287
9.2.2.1 Advantages and disadvantages of case studies 287
9.2.2.2 Sensitive evaluation 291
9.3 The questions revisited 292
9.4 Conclusions 295

NOTES

REFERENCES

APPENDICES
Appendix 1.1 Evaluation models: a selective account
Appendix 1.2 Some models of programme evaluation (Herman *et al.*)
Appendix 1.3 Dimensions of evaluation (Stake)
Appendix 1.4 Towards a conceptualisation of evaluation (Nevo)
Appendix 1.5 Evaluation studies of INSET programmes for language teachers
Appendix 2.1 Application of an evaluation checklist (Hargreaves)
Appendix 2.2 ODA Project framework
Appendix 2.3 Standards for evaluations of educational programs, projects and materials (Stufflebeam)
Appendix 3.1 Formats for programme evaluation by participants (Woodward; Vasconcelos)
Appendix 3.2 Example of Rapid Assessment Form
Appendix 3.3 Action letters (Newstrom and Scammell)
Appendix 3.4 Briefing notes for implementation plan (Alderson)
Appendix 3.5 Example of course log (Buckley and Caple)
Appendix 3.6 Student response slip (Parker and Graham)
Appendix 4.1 Letter to institutions
Appendix 4.2 Courses checklist
Appendix 4.3 Questionnaire
Appendix 4.4 Summary of evaluation procedures used
Appendix 4.5 Level of satisfaction with evaluation procedures used
Appendix 4.6 Interviews with Informants 1-5: edited transcripts
Appendix 5.1 Pre-seminar questionnaire
Appendix 5.2 Post-seminar questionnaire

Appendix 6.1 Original evaluation framework for TLE
Appendix 6.2 Criteria for the evaluation of TLE
Appendix 6.3 Extracts from participant study diary
Appendix 6.4 Integrated activity
Appendix 6.5 Example of Forum questions
Appendix 6.6 Extract from pro-forma used to elicit participant language-learning objectives
Appendix 6.7 End-of-course questionnaire
Appendix 6.8 Extract from post-course letter
Appendix 7.1 INSET for language teachers in the Lothian Region of Scotland
Appendix 7.2 Examples of Lothian Region L2 INSET questionnaires
Appendix 7.3 Course 93T130
Appendix 7.4 Keeping a journal
Appendix 7.5 Analysis of participant journals
Appendix 7.6 Questionnaire
Appendix 8.1 Letter 1
Appendix 8.2 Letter 2
Appendix 8.3 Further extracts from participant responses to follow-up
Appendix 8.4 Examples of action plans and reports on these

CHAPTER 1

EVALUATING INSET PROGRAMMES FOR SECOND LANGUAGE TEACHERS: A RATIONALE

1.1 INTRODUCTION

We live in a competitive, cost-conscious age. Within educational establishments in the UK, managers are now subject to intense pressure to demonstrate the quality of their courses, the quality of their quality assurance systems, and the overall efficiency of the operation. How good is this institution, this course ? Are resources appropriately utilised ? And what is the proof ? These are among the questions institutions are being asked to answer; and the only way of providing this information - or evidence - is through properly organised evaluation procedures.

The central question explored in this thesis is what 'properly organised evaluation procedures' might look like, and the extent to which such procedures are already standard practice.

More narrowly, the thesis is concerned with **the evaluation within UK institutions of programmes for practising language teachers**, language teachers being here taken to mean teachers of English to speakers of other languages (ESOL) and teachers of modern foreign languages (ML) to adults. To specify the field of enquiry in this way is to make two implicit distinctions, one between in-service (INSET) and pre-service (PRESET) programmes, and the other between programmes for teachers of a second or foreign language (L2) and programmes for teachers of other disciplines. Section 1.4, below, discusses these distinctions and offers a justification for the focus adopted. As a preliminary, however, I define a number of key terms (section 1.2) and offer a general rationale for programme evaluation (section 1.3).

1.2 THE NATURE OF PROGRAMME EVALUATION

1.2.2 Programmes, projects and courses

Contrasting programmes and projects, Weir and Roberts (1994: 3) define *projects* as 'educational activities ... subjected to contractual definition and finite time-scales' and 'funded to achieve a particular task'; *programmes* are more loosely defined as 'any organized educational activity offered on a continuing basis' (*ibid.*). A programme might thus be 'an innovative syllabus, a teacher-upgrading course, or any of a variety of applied linguistics related activities' *(ibid.*). A similarly broad definition is offered in Herman, Morris and Fitz-Gibbon's (1987) 'Evaluator's Handbook', which forms the foundation for the highly respected Sage Program Evaluation Kit (various authors, 1987). 'A program,' say Herman *et al* (1987: 8), 'is anything you try because you think it will have an effect.' They continue, more concretely, but no more persuasively: 'A program might be something tangible, such as a set of curriculum materials; a procedure, such as the distribution of financial aid; or an arrangement of roles and responsibilities, such as the reshuffling of administrative staff...' (*ibid.*). This is perhaps less strange than it seems at first sight, since all of these examples, like Weir and Roberts' syllabuses, while falling far short of any normal definition of 'programme' within educational settings, might form appropriate foci within programme evaluation.

For Weir and Roberts (*op.cit.*), following Joint Committee (1981), the key distinction between programme and project lies in the expected lifespan, i.e. the fact that project funding is limited-term. While this may be one way of distinguishing between a programme and the type of project which is no more than a course, many projects are much more complex in terms of their scope (see e.g. Kadepurkar 1997) and intended effects (Chambers and Erith 1990), and are therefore much more difficult to evaluate for this reason. A more important distinction might therefore be that programmes are 'continuing' (Weir and Roberts, *op.cit.*).

I shall be adopting what I take to be a traditional view of programme, as an instructional plan or set of structured learning activities which is limited by time (and often space) and *which is repeated or potentially repeatable*. Reference will be made to projects only when this seems relevant to the discussion of programmes. In a choice between 'programme' and

'course', 'programme' will be the unmarked form and 'course' will be used either to refer to a unique event or to a single instance within a series of similar events (the programme). This distinction not only has practical implications, in that programme evaluation, unlike course evaluation, is likely to have a developmental aspect, it also has a pragmatic justification. Whereas managers of institutions might be less inclined to spend money on the evaluation of one-off events (courses), they may well see the benefits of evaluating a *programme* in the interests of efficiency or in an attempt to achieve a higher level of client satisfaction.

1.2.2 Evaluation

In its early days educational evaluation was associated with measurement and learner assessment (see Appendix 1.1 for a brief account of the development of the discipline); these days it is a much broader concept and may be as much concerned with information gathering on processes as with judgements on outcomes (McGinley 1986; Sharp 1990; Rea-Dickins and Germaine 1992; Alderson 1992; Rea-Dickins 1994; Weir and Roberts 1994; Murphy 1995). Tests, it is generally agreed, are appropriate as a procedure within evaluation when quantified evidence is required of the gains made as a result of certain kinds of programme. However, programmes cannot be evaluated purely on the basis of test results (Cumaranatunge 1989).

For some writers (e.g. Elliott 1980), evaluation is equated with research, or the differences between evaluation and research are seen as less important than the similarities (Beretta 1990a). The most widely held view, however, seems to be that evaluation and research converge and overlap at a procedural level but diverge in relation to conception and purpose (Isaac and Michael 1981). Evaluation can be defined (after Scriven 1996) as the study of quality (or effectiveness) and value (or efficiency) and is normally perceived as being oriented towards practical decision-making (Rudduck, R. 1981). This applied dimension to evaluation, and such considerations as time and resource constraints and 'the demands of policy makers for "answers" and "results"' (Rea-Dickins 1994: 72), renders it distinct from basic research; it may nevertheless be seen as a specific form of applied research (Patton 1990; Nunan 1992) if it satisfies normal scientific criteria relating to data collection and analysis (e.g. systematicity, reliability, internal validity). Rea-Dickins (1994), who provides a helpful summary of the various views on the relationship between

research and evaluation in the general education and applied linguistics literature (the latter being very limited), suggests that research and evaluation are best thought of as being part of a continuum, with evaluation as an 'extension' of research (Rea-Dickins 1994: 73). For the purposes of this thesis, evaluation will be treated as a form of (applied) research activity concerned with the determination of quality and value and characterised - as far as is practicable within real-world constraints - by the features of what Beretta calls 'disciplined inquiry' (1990a: 11).

1.2.3 Programmes and their evaluation

A programme may be perceived as a single entity but it is certainly not conceived in this way. Any programme development model these days would make reference to at least three stages. Stage 1 is the fact-finding stage, in which there is some consideration of contextual constraints. Stage 2 is the design stage, in which decisions are taken on aims and objectives, content, sequencing and timetabling, materials and resources, and delivery. And stage 3, following on naturally from these planning stages, is the implementation stage.

Each of these three stages is potentially susceptible to evaluation. In relation to stage 1, we might evaluate for instance the adequacy of needs analysis procedures; the degree of consultation with interested parties ('stakeholders'), the involvement of whom might affect the use made of results; or the planning of 'baseline' studies to establish the *status quo*, if these were conducted. The conceptual framework of stage 1 can be assessed at the time, by a process of 'armchair evaluation', but it will eventually be judged by its outcomes - the extent to which the resulting information is available, correct and useful. During stage 2, there is likely to be continuous evaluative discussion around a series of proposals and revisions (Cumaranatunge 1989). Such discussions will be concerned with, for example, the relevance or realism of course aims, the appropriateness of content, and resource implications. It is stage 3, however, that tends to be the focus of most overt, reported evaluative activity, both ongoing and *post hoc*. This is the point at which a range of decisions relating to the selection, sequencing and balance of components will be tested, but it is also a time when there will be evaluation of elements, such as tutors' presentation skills, efficiency of administration or participants' judgements of value for money, which are less amenable to pre-planning because they depend on the quality of delivery.

One of the problems for evaluation is the discreteness of the elements within these stages. The next chapter takes up this issue with regard to such questions as the level of evaluation foci, the selection of foci, and the means by which foci are selected; at this point, the intention is simply to draw attention to the fact that the evaluation of a programme necessarily involves the evaluation of separate components. Where an evaluation is concerned with specific aspects of a programme, as in formative evaluation, this is not a problem, but where a wide-ranging summative assessment is required, it may well be. Even if it were possible to evaluate 'everything', a programme will always be more than the sum of its parts and in order to make global statements integrative, interpretive judgements are needed. These judgements may be based on quantitative evidence and/or considerable experience, but they are no less subjective for that. We return to this issue in later chapters.

A second problem is that in many cases evaluation which stops at the programme implementation stage (stage 3) will be incomplete because it fails to take account of post-programme effects. The importance of post-programme evaluation and the difficulties involved is a major focus of the thesis.

1.3 THE NEED FOR PROGRAMME EVALUATION

Various reasons have been advanced for the importance of programme evaluation. These can be summarised as follows:

ñ evaluation is a source of information on general issues to do with the facilitation of learning;

ñ evaluation provides an informed basis for continuation or change;

ñ evaluation is a key factor in good management;

ñ evaluation is a stimulus to personal development.

The most obvious general reason for evaluation is that it provides what Weir and Roberts call 'professional information' (1994: 11). This, they suggest, can be used to settle theoretical questions about suitable directions to be followed or it can take the form of context-sensitive information on what works under given circumstances. The two are

potentially related, since an understanding of why given procedures, say, work more or less well - if the evaluation yields insights of this kind - can make an important contribution to decisions regarding future directions in a specific context elsewhere (Cumming 1988; Weir and Roberts 1994); and this indirect effect can be even greater than that felt locally (Mitchell 1991).

Henderson (1980), describing an Open University programme, distinguishes between 'process' evaluations, which are intended to 'obtain information on the success or otherwise of the course as a learning experience' and 'product' evaluations, which are designed to 'discover the impact of the course on teachers' school practices' (1980: 70). Although one Henderson's definition of 'process' is arguably too narrow, the distinction between process and product is important: we need to know what went on during a course in order to check that what was intended to happen did happen but also to understand why the outcomes were what they were. As noted above, such information is potentially valuable not only to those within the institution where the programme is offered.

Within institutions, evaluation serves the needs of programme organisers and managers, validating certain arrangements and practices, identifying areas of weakness and possibly highlighting constraints on implementation - for Rudduck, J. '*the* critical in-service problem' (1981: 146, original emphasis). It thus provides a basis for planned change (Rea-Dickins and Germaine 1992). The records and data generated by evaluation are also a valuable source of information for others involved in the programme subsequently and for programme review; for this purpose, it is helpful if a note is kept of follow-up action taken (Weir and Roberts 1994).

The responsibility for follow-up action and, indeed, for programme evaluation in institutions in general, rests with managers. Managers may in practice delegate certain evaluation tasks to teaching staff but the allocation of the necessary resources, including time, is a function of management. Evaluation is part of the cycle of good management (Mackay 1994a), which 'integrates both continuous and final evaluation into any educational and teaching plan' (Rea-Dickins and Germaine 1992: 14). Since institutions are after all 'prone to routine, inadequacy and ineffectiveness', to assess what they are doing on a regular basis will prompt further development on both an institutional and a

personal level (Cumming 1988: 44). Where consciousness is created of problems relating to personnel, this might have a variety of consequences, ranging from a recognition of the need for greater support or further training to contract termination. In short, evaluation is a key input to decisions on pedagogical issues, personnel and resources, where these are shown to be inadequate. More positively, it can justify the use of resources, and thereby be a means of retaining support for publicly-funded programmes (Ashworth 1985).

The personal effects of evaluation are manifold. Programme personnel develop professionally from the scrutiny of their performance, especially if this is carried out in a supportive atmosphere and forms part of an appraisal procedure. If they have an opportunity to work with an outside evaluator they may also learn from this collaboration. Moreover, participants who experience non-threatening evaluation may come to feel less suspicious of evaluation than might otherwise be the case and go on to make use of similar procedures in their own classrooms. The nature of the interaction between programme personnel and participants involved in some forms of evaluation (e.g. dialogue journals) may also have a positive effect on their relationship. Murphy's (1995, 1996) study of trainee evaluators enumerates a number of other self-reported benefits (e.g. broader perspective on teaching; clearer sense of what needs to be changed).

Programme evaluation thus benefits not simply the programme itself or future participants but conceivably also programme personnel, the institution, and the professional community at large.

1.4 THE SPECIFICITY OF INSET FOR LANGUAGE TEACHERS

1.4.1 Introduction

As indicated in Section 1.1, the focus of the thesis is the evaluation of UK-based INSET programmes for L2 teachers. The exclusions implicit in that focus are illustrated in graphic form in Figure 1, below:

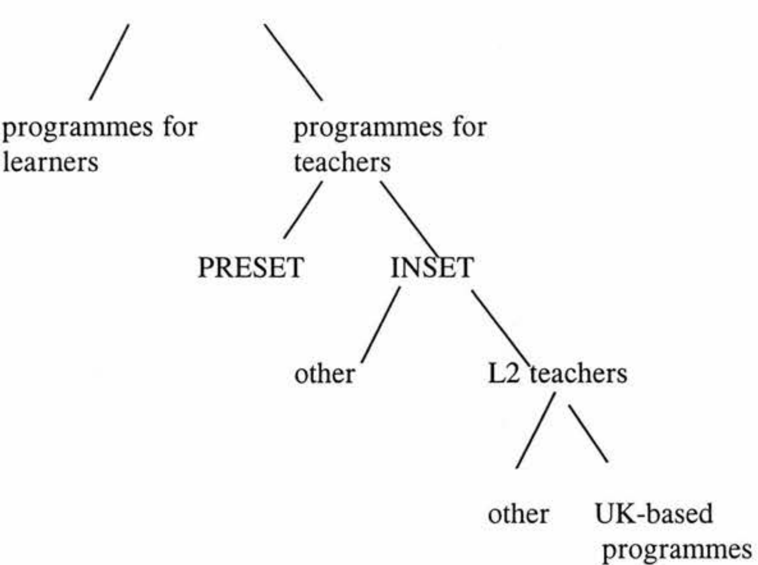


Figure 1: Focussing down

Since there exists an underpinning discipline, (programme) evaluation, there will logically be many points of overlap at the level of principle and practice between the types of programme within each dichotomy represented in the diagram, and to a lesser extent across dichotomies, with programmes for learners (for whom language is primarily a subject of study) being the obvious exception. However, the argument that will be presented is that differences in the participants, aims, content, organisation and context of the various programme types necessitate different evaluation designs, and that the role of language in programmes for practising language teachers, and the possible location of such courses, may be a reason for treating these as a special subcategory within the broader category of INSET programmes.

Table 1 below offers a characterisation of programmes for language teachers. These are distinguished according to such features as the level of programme (PRESET/INSET), whether it is award bearing, and its location (in-country or not). Some of the points of difference which have particular implications for programme evaluation are discussed below.

Table 1: Types of programme for language teachers

	PRESET 1	PRESET 2	PRESET 3	INSET 1	INSET 2	INSET 3	INSET 4
PURPOSE	meet demand for qualified teachers	increase supply by 'converting' qualified teachers	meet demand from prospective teachers for initial qualification	introduce innovation	develop cadre of specialists	provide for continuing professional development	meet demand from teachers for higher-level qualification
AWARD BEARING	Yes	Yes	Yes	No	Perhaps	No	Yes
PARTICIPANT	inexperienced NNS	qualified teachers of other subject (NNS)	NS (inexperienced or qualified teacher of other subject)	practising teachers (NNS)	practising teachers (NNS)	practising teachers (NS/NNS)	practising teachers (NS/NNS)
SCALE	large	small	large	large	small	large	small
FUNDING	state	state	private	state	state	state/private	state/private
ATTENDANCE	compulsory	optional	'optional'	compulsory	optional	optional	optional
LENGTH	1-5 years	1 year part-time	4 weeks	very limited	up to 1 year	short	up to 1 year
LOCATION	in-country	in-country	in-country	in-country	for NNS, overseas?	for NNS, in-country or overseas	for NNS, overseas?

1.4.2 INSET evaluation *vs* PRESET evaluation

From a narrow perspective, PRESET programmes are designed to produce teachers with a level of subject specific and general pedagogic competence such that they can perform their teaching duties in a manner appropriate to their status as beginner teachers. A broader view might emphasise the importance of teachers having some awareness of the historical evolution of the educational system, of the philosophy of education and, in many contexts, of aspects of the native language and culture. This distinction between the narrow view and the broad view can be related to the length of the programme (*cf* the differences between PRESET programme types 1, 2 and 3 in Table 1), but also to the view held within a particular educational system or organisation of the purpose of teacher education (the training/education debate).

The underlying vision of programme purpose (crudely expressed by the terms training and education) has implications for programme evaluation (e.g. what is evaluated and when) which are equally relevant in a general sense to INSET programmes. However, there are a number of important differences between the two levels of programme which influence the way in which they are normally evaluated.

There is the issue of programme aims, for instance. Whereas INSET programmes tend to take a foundation of knowledge and skill for granted and build on it, PRESET programmes seek to establish this foundation; in the case of language teachers, this foundation consists of linguistic as well as pedagogic knowledge and skill. Because (with the exception of a tiny proportion of 'taster' courses) PRESET programmes are also award bearing, evaluation tends to concentrate on the extent to which participants have achieved criterial levels in key curriculum areas and is limited to performance within the duration of the course (i.e. before participants are certified fit to teach).[1] It is hardly surprising, then, if the success of a programme is judged by the number of passes. Where there is a need to satisfy a quota, this may be the only measure accorded any importance by the bureaucracy.

Although pass rates might similarly be adduced as a criterion for programme success in relation to award bearing INSET programmes, other criteria are arguably more relevant. Most INSET activity - and this includes award bearing programmes - is less concerned

with the acquisition of subject knowledge than with awareness-raising and attitude-change in relation to new ways of doing or thinking. The real test of whether a programme has been effective in these terms is whether the awareness is subsequently translated into changed practice. Moreover, there is a further reason why the effectiveness of INSET can only be evaluated when participants are back in their own teaching contexts. As practising teachers, they will be in a position to judge the relevance of content, the quality of delivery and materials or the adequacy of administrative arrangements while the course is in progress, and course organisers sensitive to participants' status as co-professionals will want to know their views on such features. However, the workability of ideas can only be judged when participants have an opportunity to try them out in their own classrooms. In relation to the nature of effects, then, and the timing of the evaluation of these, an INSET programme for L2 teachers has more in common with INSET programmes for teachers of other subjects than with subject-specific PRESET.

1.4.3 INSET programmes for language teachers

INSET programmes for language teachers also differ in a number of respects from programmes for teachers of other subjects. These differences are all the more striking in the case of L2 teachers who are not native speakers of the languages they teach.

The most obvious difference is in relation to the role played by language. Whereas the target language (TL) will constitute a major (perhaps *the* major) component in PRESET programmes for L2 teachers, in the forms of language development and language awareness (the latter also a major component in programmes for native speaker (NS) L2 teachers), in INSET programmes it will seldom be the principal component and may not even be an overt focus. The relative brevity of many INSET programmes is one of the factors that lies behind this shift of focus, others being changed priorities (on the part of programme organisers) and perceived need (on the part of participants). Nevertheless, the use of the TL as the medium of instruction on programmes for non-native speakers (NNSs) undoubtedly serves the secondary purpose of providing at least minimally for the maintenance of participants' linguistic skills and optimally for their further development. The effect of TL-use during such programmes might indeed be an appropriate focus for evaluation.

A further difference between programmes for L2 teachers (and this applies in different ways to NSs and NNSs) and those for teachers of other subjects is that the former may take place in a different country from that in which participants teach. The NS TESOL teacher might do a course in the UK, but then teach overseas. For the NNS TESOL teacher, there are obvious benefits to be derived from time spent in an English-speaking country, and from a stay in which language exposure is combined with professional training (see INSET programme types 2, 3 and 4 in Table 1.1). The evaluation of programmes for both types of participant, NS or NNS, poses problems beyond those that might normally be encountered in relation to an in-country programme. Probably the most significant of these problems are *the increased difficulty* - due to the distance and dispersion of participants, even those from the same country - *and expense of collecting baseline data and data on post-programme effects*. When the overseas programme forms part of a large-scale project involving the introduction of a new syllabus and/or new materials and/or a new examination system, or when it is part of a larger sandwich, with other phases taking place in-country, there will be even greater problems.

1.4.4 Summary

This section has suggested that in evaluation terms programmes for practising L2 teachers are a special case but also that they have more in common with other INSET than with programmes for trainee language teachers.

PRESET programmes will be excluded from further discussion. In view of the difficulty of obtaining information on INSET evaluation practices outside the UK, INSET programme Type 1 also falls outside the scope of the discussion, although some of the recommendations made in the final chapter may be judged to have some relevance for these contexts too. In subsequent sections, we take a brief look first at the literature concerned with the evaluation of INSET programmes for UK schoolteachers, and then at work on programme evaluation within language teaching and language teacher education.

1.5 INSET EVALUATION IN THE UK: A BRIEF HISTORY

For the purposes of the present work INSET is taken to be the provision within a framework of formally organised activities of opportunities for the professional

development of serving teachers. This is not to deny that many of the most valuable forms of professional development will be self-initiated or even that such initiatives may be set in motion by external events. The effect, however, is to exclude consideration of action research and teacher self-evaluation where these take place as self directed activities.

Writing in 1981 and referring to LEA provision within the UK, Sparrow claims that 'short courses are and always have been the most widely-used form of in-service training' (Foreword to Rudduck, J. 1981: 10). Rudduck goes on to say that the short course is 'only one of a rich array of in-service possibilities' (1981: 17). She is here thinking also of the kinds of informal activity in which teachers might be involved (teachers' panels, committees, study groups, individual study) which might lead to teacher development. Within LEAs these days formal provision is itself more flexible. Rather than centralised courses in which the agenda is set by the LEA, there has been a move towards school-focused and school-based activity (the main difference being in the degree of specificity), for which the starting-point are the contextually-defined needs of the participating schools or school. One of the benefits of this changed situation is that provision is more likely to be relevant; possible disadvantages include the fact that very similar forms of training may need to be offered on a number of occasions, thus adding to the workload of those providing the training.

The primary concern here is not the education and training of state-school language teachers within LEAs (although reference is made in Chapter 6 to a specific example of such provision within one of the Scottish regions) but with the kinds of programme provided by UK institutions such as universities and private language schools for ESOL teachers or teachers of ML to adults. However, since UK institutions, whether schools on the one hand or universities, colleges and private language schools on the other, operate within the same broad educational context, it seems relevant to make brief reference to the development of INSET evaluation in relation to the state school sector.[2]

As noted earlier, one of the functions of evaluation is to provide information to guide decisions concerning future provision. With hindsight, it seems rather strange that a recognition of the necessity for such information has only manifested itself within British educational circles within the last 25 years or so (see Appendix 1.1 for the historical context). Rudduck, J. (1981) refers to the report of a conference on INSET in 1972, the

year of the James Report. The conference report includes a statement by Lord James in which he draws attention to the lack of information about the potential effects of INSET:

> [we do not know] what effect various kinds of postexperience training actually have on teaching and the teacher; how long these effects last; what are the most appropriate kinds of education to accomplish ends which may be quite different for different individuals; and what effects on the schools
>
> themselves the in-service education of their staff has.
>
> (Watkins 1972: 12, cited in Rudduck 1981: 15)

Stenhouse, writing just a few years after this conference, points out that educational evaluation, although 'highly developed' in the USA and Sweden, is only just beginning in the UK (1975: 98). It is certainly true that educational evaluation was already well established in the USA by that point, but the focus of this activity was not teacher education. Tisher and Wideen (1990), editors of an international collection of survey papers (Wideen and Tisher 1990a), commenting on the fact that little or no research is reported into teacher education in the two decades after the Second World War, infer that the priority in that period was the development of programmes rather than research: 'Teacher education had not yet reached the stage of academic self-consciousness where research was looked at as a necessary endeavour' (Tisher and Wideen: 1990: 5). One of the specific lacunae noted by contributors to Wideen and Tisher (*op.cit.*) and singled out for comment in the editors' concluding chapter (Wideen and Tisher 1990b) is the lack of systematic evaluation studies.

Some hard data is available on this point for the UK. Henderson (1978) refers to a study by Henderson, Perry and Spencer (1975) which surveyed 1,044 INSET events over the period 1968-71 in one area of the UK. Only 31 of these events appear to have been evaluated in any way. Of these, 24 were award bearing courses in which evaluation was interpreted to mean assessment of participants; the other seven, all reported by the same organiser, made use of a short end-of-course questionnaire designed for programme development purposes. Henderson, whose interest in evaluation can be explained by his asociation with the Open University, concludes: 'There is no evidence that the problem of evaluation had been tackled any more seriously elsewhere in the UK at that time' (1978: 44).[3]

The later part of the 1970s were different, however, in that programme evaluation became a specific focus. Some evaluation studies are reported in Henderson (1978) and it is probably no coincidence that four books on INSET, three of which were exclusively concerned with evaluation, were published around the turn of the decade. Fox (1980) is a monograph arising from a series of conferences on educational evaluation held in the second half of the 1970s and case studies from a number of countries, including the UK. McCabe (1980a) is a collection of papers - many of them case studies - by UK evaluators; it includes a survey of UK evaluation studies in the 1970s (Taylor 1980). Salmon (1981) is a report of a workshop attended by evaluators from the Council of Europe. Rudduck, J. (1981) is a full-length study of the contribution to INSET of the short course; the chapter on evaluation draws heavily on the work of two Suffolk advisers, Saville and Andrews.

The 1970s had been a period of expansion in Europe as far as teacher recruitment was concerned; falling school rolls in the 1980s meant a decline in the demand for new teachers and therefore a more significant role for INSET, as the most obvious means of introducing new ideas into classrooms (Vorbeck 1981). Curiously, the publication of the volumes referred to above seems not to have resulted in a significant increase in reported evaluation studies. One reason may have been lack of general awareness of these publications. Hodgson and Whalley (1985: 44), for instance, point out that 'a respectable range of literature' has been produced on the evaluation of pre-service courses, but 'the evaluation of in-service work, particularly of complete courses, has received rather less attention, despite the rapid expansion of such provision over the past two decades'. They mention the work of Henderson (1976, 1978), but surprisingly make no reference to such publications as McCabe (*op.cit.*) or Rudduck, J. (*op.cit.)*. Walker, also writing in 1985, while acknowledging the work that had been done in educational evaluation, notes: 'a significant gap which remains is in the area of research methods and techniques' (1985: 2); given what has been said earlier concerning the relationship between research and evaluation, it is reasonable to see this as a comment on methods as tools of both research and evaluation. Although a number of works on research methods have since been published, so far as I can ascertain no major work making the link to INSET evaluation has appeared to fill that gap.

1.6 EVALUATION IN SECOND LANGUAGE EDUCATION

1.6.1 Introduction

The distinction made earlier in this chapter between programmes for second language learners and programmes for teachers does not hold good for the personnel involved in programme design, implementation or evaluation. Nor does it hold good for those who write about programme evaluation, many of whom have moved on from working with learners to working with teachers, while others continue to combine the two. The field within which we all work is second language education. For this reason, it seems appropriate to preface an overview of the literature on the evaluation of programmes for second language teachers with an indication of work on the evaluation of programmes for learners.

1.6.2 The evaluation of programmes for language learners

Three recent papers provide, individually and collectively, a comprehensive picture of evaluation activity in language teaching over the last twenty years (Mitchell 1991; Beretta 1992; Rea-Dickins 1994).

Beretta's (1992) review is uncompromisingly critical, the particular object of his criticism being the 'methods studies', both large-scale (Keating 1963; Scherer and Wertheimer 1964; Smith 1970), and small scale (the paper includes a summary of 33 studies carried out between 1963 and 1985). Although the work of Mackay (Mackay 1981) is specifically exempted from the general criticism, Beretta's slightly hedged conclusion is that 'Probably none of the studies serves as a particularly useful guide to evaluators of language education programs today' (1992: 12). Indeed Beretta (1990b) wonders, as had Mackay (1981) before him, whether the lack of success of the methods studies had not for some time acted as a deterrent to further evaluation activity.

Like Beretta, Mitchell (1991) comments critically on the methods studies and draws a comparison between developments in general educational evaluation and the literature of the late 1970s and 1980s on the evaluation of language programmes, finding within the latter 'strikingly uneven levels of awareness of the debates within mainstream educational research' (1991: 148) and a lack of explicitness concerning the rationale for the choice of

evaluation procedures. She also draws attention to the quantitative orientation of the bulk of the studies during this period (a small number of Canadian studies being the exception to this general rule), arguing instead for a 'multifaceted approach' as exemplified in her own study of bilingual education in the Western Isles.

The value of the papers by Beretta (1992) and Mitchell lies partly in their breadth, the fact that they take in work in TML as well as TESOL, but also in their critical edge. However, this edge is also their limitation as surveys, since each has an axe to grind. Rea-Dickins's (1994) survey is narrower in that it deals only with TESOL. She makes similar criticisms to those of Beretta and Mitchell concerning the insularity of TESOL, but is more concerned to describe trends (within evaluation in general and not just programme evaluation) than to evaluate specific features of the work surveyed. The generally pessimistic tone of the review provides a useful point of comparison with comments made some years earlier by Perkins and Angelis (1985), McGinley (1986) and Alderson (1985a).

Each of the above had referred to the relative paucity of evaluation activity within applied linguistics and (English) language teaching, but had sounded a hopeful note. For instance, the abstract of Perkins and Angelis's paper on language programme evaluation in the *RELC Journal* contains the following statement:

> There is a robust tradition of program evaluation in the field of testing and measurement, but within the TESOL community, program evaluation has only recently begun to be considered as a vital component.
>
> (Perkins and Angelis 1985: 72)

Since this is the only paper on the topic to appear in the journal during the period 1980-1995, it clearly had little effect in terms of stimulating academic debate among the readers of the journal. The same might be said of McGinley's (1986) paper in *System*, which sets comments on what McGinley sees as the neglect of evaluation in English for Specific Purposes (ESP), 'both in the literature and in practice' (1986: 335) in the context of a thumbnail sketch of the general evaluation literature. McGinley offers what he calls a speculative framework for the evaluation of ESP programmes in the hope that this will lead to more systematic and open scrutiny of the processes involved. Judging by subsequent issues of the same journal, his proposal fell on deaf ears.

McGinley is perhaps unduly negative. While it is true that publications concerned with the evaluation of English language teaching programmes in general had been few in number (see Rea-Dickins's 1994 survey review), there was evidence of a nascent literature relating to the evaluation of (SP) programmes and pioneered by Mackay [4] (Mackay 1981; a paper by Bachman (1981) in Mackay and Palmer 1981; Alderson and Waters 1982; Kennedy 1985). The paper by Alderson and Waters, which described a course in testing and evaluation for language teachers, was a little different in focus from the others mentioned but like them was ultimately concerned with the evaluation of ESP programmes.

The third statement appears in Alderson (1985a):

> Until recently, not a great deal of attention was paid in language teaching, especially the learning of English as a Foreign Language, to the evaluation of the course's success.... Teachers would equate success with their own survival ... by attendance rates, or by the warmth of the after-glow at course end: the bigger the buzz, the more successful the course/materials/ teachers. Warmth, friendliness, niceness, atmosphere, such were the criteria, usually implicit, for judging courses. There were, of course, honourable and important exceptions to this. Today, however, there is an increasing understanding of the importance of objectifiable evaluation of courses and the methodology, content and interpretation of curriculum evaluation is becoming developed and understood. Even in applied linguistics and language teaching, the central importance of evaluation is being recognized and publications in this area are increasing in number.
>
> (Alderson 1985a: 129).

On the evidence of publications alone the confidence evident in the last sentence of this quotation seems a little misplaced, all the more so since Alderson's paper was originally conceived some years earlier than it appeared (Alderson, personal communication). The published literature may of course present a very imperfect impression of actual activity. Rea-Dickins (1994: 73) points out that 'one explanation for the apparent dearth of evaluation studies is that evaluative data may not be disseminated beyond the immediate sponsors, institution, or programme' and that reporting may be in an informal and unwritten format. While this is probably the normal case, there are instances of unpublished work being fairly widely disseminated. Writing of the early 1970s in Britain, Walker (1985) refers to the circulation of an 'underground literature' - coincidentally concerned with evaluation - mainly in the form of case studies. These unpublished papers

constituted, he says, a kind of 'invisible college' (1985: 2)[5]. What Alderson had in mind apparently (personal communication), was not so much books or journal publications but less visible activities such as conference papers, discussions at seminars, and the reports, only a small number of which were published (British Council 1981; Rea 1983; ODA 1984), resulting from a growing recognition on the part of aid agencies that evaluation was necessary. Although a number of key figures such as Alderson were clearly aware of developments in educational evaluation, the key factors in the gradual growth of a field-specific evaluation literature appear to have been external: the trend towards projectisation in ELT aid programmes, a suggestion also made by Rea-Dickins (1994) and, *pace* McGinley, the demand for accountability within SP programmes (Mitchell 1991).

Despite the little clusters of work in SP and project evaluation, there is a feeling of miscellaneity about the publications referred to above. These were the product of independent voices rather than the massing of a choir. From the perspective of this thesis, the publications were also limited in two important respects:

1. With a small number of exceptions, as noted above, they showed little awareness of important work done earlier within the field of educational evaluation. One interesting explanation for this is that applied linguists have tended to be located in departments of linguistics rather than departments of education and have therefore been isolated to some extent from the general debate concerning the evaluation of school curriculum projects (Alderson, personal communication).

2. Their main focus was language teaching. Few references can be found to the evaluation of teacher education programmes, despite the fact that, as indicated above, the effectiveness of in-service teacher education was the subject of intense debate in the late 1970s and early 1980s. In this respect, Alderson's paper (1985a), which discusses the difficulties of post-course evaluation and makes specific reference to INSET programmes, is of particular interest.

The ironic note in the final sentence of the above quotation from Alderson's paper ('Even in applied linguistics and language teaching ...') can be explained by the first of the limitations noted above - until recently TESOL at least has certainly tended to be rather narrow in its preoccupations and has failed to draw on ideas from other fields. What is

equally ironic, however, with hindsight, is the fact that the publication of 'Evaluation', the volume edited by Alderson (1985b) in which his paper appeared, was probably quite widely read within a narrow circle, and yet, like the papers cited above by Perkins and Angelis and McGinley, seems to have had little immediate effect in terms of stimulating further public debate about programme evaluation.

Between 1986 and 1990 the trickle of papers continued. There were, for instance, more papers by Rea-Dickins (1987) and Mackay (1988); a number of wide-ranging articles by Beretta (1986a, 1986b, 1986c, 1990a, 1990b); a flurry of papers on the evaluation of the Bangalore project (Brumfit 1984; Beretta and Davies 1985; Greenwood 1985); and a section of Johnson (1989) was also devoted to evaluation. These papers apart, the topic of evaluation receives relatively little attention in the applied linguistics/language teaching literature until the 1990s, when there is an explosion of activity: book-length publications (Anivan 1991; Alderson and Beretta 1992, which contains the survey referred to above by Beretta (1992)[6]; Rea-Dickins and Germaine 1992; Weir and Roberts 1994; Rea-Dickins and Lwaitama 1995); a newsletter (PRODESS News) oriented towards evaluation, the product of a joint venture between Thames Valley University and the British Council and aimed principally at project staff working in Central and Eastern Europe; a state-of-the-art article (Rea-Dickins 1994), and professional meetings devoted to evaluation (e.g. annual TESOL America colloquia on evaluation from 1992; British Council 1995).[7].

The upsurge in interest in evaluation confidently predicted by Alderson a decade or more ago has finally happened, then. Or has it ? In 1992 Beretta (*op. cit.*) noted sombrely that there is no evidence that evaluation is being taken seriously within the field of second language education as a whole, and more recently Potter (1994), Murphy (1994) and, as noted above, Rea-Dickins (1994) have echoed these sentiments. Are the recent publications, contrary to these pronouncements, evidence of widespread interest and awareness at practitioner level, a public manifestation of what has until recently been lively activity beneath the surface or - since all the authors/editors are associated with universities - testimony to a major new development pioneered by a select group who are still years ahead of the rest of the profession ? A partial answer to that question will be offered in Chapter 4.

1.6.3 The evaluation of programmes for language teachers

At the more specific level, the evaluation of programmes for second language teachers, there is little reported activity. Since the late 1960s there has been an increasing supply of books and journals for language teachers, but it was only in 1987 that the specific needs and concerns of language teacher educators found a focus in the journal *The Teacher Trainer*; and several more years passed before the first books for language teacher educators became available through mainstream publishers (e.g. Doff 1990; Richards and Nunan 1990; Wallace 1991; Woodward 1991, 1992)[8]. Documenting the growth of this literature, Rea-Dickins (1994) finds little evidence that 'assumed good practice' is being evaluated 'in any systematic way' (Rea Dickins 1994: 81). In fact, with the exception of Woodward (1991), none of the latter publications has more than a passing reference to the evaluation of language teacher education programmes[9].

Within the field-specific evaluation literature already referred to, Weir and Roberts (1994) contains case-studies of an INSET programme (Nepal) and a PRESET programme (Paraguay) and there are references to teacher education in two of the case studies (Alderson and Scott, Coleman) in Alderson and Beretta (1992) and in the PRODESS Newsletters and Colloquium Papers (Kiely, Murphy and Rea-Dickins 1994; British Council 1995). For the rest, a small number of publications present isolated evaluation studies of INSET programmes. These studies, only six of which relate to programmes based in UK institutions, are summarised in Appendix 1.5 and referred to in Chapters 2 and 3.

To echo Rea-Dickins (1994), there is clearly a need for studies which evaluate recent approaches to teacher education and thereafter disseminate good practice. The external pressures referred to in the opening paragraph of this thesis, together with an increasing emphasis on 'reflective practice', should lead to an increase in this kind of publication, but to judge from the studies listed in Appendix 1.5, there is a need for wide-ranging debate on such questions as criteria for evaluation, the choice of evaluator, and the timing of evaluation. Given the discussion in section 1.4 of the difficulties associated with the evaluation of specific types of INSET programme, there is also a need for special consideration to be given to evaluation method.

1.7 RESEARCH QUESTIONS, METHOD AND STRUCTURE OF THE THESIS

It has been suggested in relation to programmes for learners that non-specific educational evaluation is well in advance of discipline-specific developments (Beretta 1992). To judge from the published literature, this is just as true for the evaluation of INSET programmes.

There is an urgent need for two kinds of evaluation-related research in the field of L2 teacher education. These correspond roughly to what Scriven (1996) has termed 'meta-analysis' (or research synthesis) and 'meta-evaluation' (the evaluation of evaluation method). Such research would include:

1. the gathering and dissemination of information through mainstream channels: i.e. not simply the bringing together of reported practices and the rationale for these but also the externalisation of hitherto unreported practices;

2. awareness-raising in relation to those evaluation methods (or possible methods) which do not appear to be widely used within INSET and active experimentation with a view to establishing their potential contribution.

For the purposes of the present work, these concerns have been translated into the following research questions:

1. *What methods and techniques are available for the evaluation of INSET programmes for language teachers ?*

2. *What is the situation within UK institutions with regard to the evaluation of INSET programmes for language teachers ? How are such programmes evaluated? Are there any patterns in the procedures used or not used ? How does actual practice compare with that recommended in the evaluation literature ? How aware are those responsible for evaluation of any inadequacies in approaches to evaluation within their institution ?*

3. *Of the procedures that do not appear to be commonly used, which might make a*

positive difference to the effectiveness or efficiency of evaluation ?

These questions are answered by means of a literature review, a survey and a series of case studies.

The literature review, briefly introduced in previous sections of this chapter, is extended in Chapters 2 and 3. In Chapter 2, a 'decision-making hierarchy' is proposed. As the label suggests, this model is intended to raise awareness of the interrelatedness of the various decisions taken during the planning and implementation stages of evaluation, and thereby raise discussion to a level of abstraction above that of the isolated *Wh-* questions favoured by many writers; it also serves as an organisational basis for the chapter as a whole. Chapter 3 then takes a close look at one of the points in the hierarchy to which most decisions lead: evaluation method; and, using criteria suggested in the previous chapter, considers the advantages and disadvantages of specific methods and techniques. The review thus provides a partial answer to research question 1 (*What methods and techniques are available for the evaluation of INSET programmes for language teachers ?);* the answers found there are then reassessed in the light of the survey and the case studies in later chapters.

One approach to research question 2 (which concerns evaluation practices within UK institutions) might have been through published accounts of relevant evaluation practice. Such accounts are however extremely limited in number (see Appendix 1.5). A survey was therefore conducted into the procedures used by a range of UK institutions to evaluate their INSET programmes for language teachers and the views of authoritative sources within the institutions on the appropriateness of these procedures. The results of the survey, which used postal questionnaires and interviews, are considered both normatively and against the broader context of the literature review. The survey is reported in Chapter 4.

Suggestive answers to research question 3 can be found in some of the responses to the survey, and in possibilities indicated by the literature survey but not echoed in the survey of current practice. The appropriateness of some of these ideas is explored by means of a series of case studies in Chapters 5-8.

The case studies all relate to in-service provision for L2 teachers (ESOL and ML), but they vary according to focus, scale (duration) and the nature and extent of my own involvement. The intention was to extend the discussion in Chapter 3 from a practical basis.

Case Study 1 is a small-scale study of a single seminar within a teacher development (TD) programme for teachers of modern languages at the Institute for Applied Language Studies (IALS) at the University of Edinburgh. I led the seminar and conducted the evaluation. The study highlights the value of baseline data collection and the subsequent evaluation of effects.

Case Study 2 is a longitudinal study of the evaluation procedures used on a three-week refresher course for non-native speaker ESOL teachers. The study illustrates the variety of measures that might be used for the formative and summative evaluation of such programmes and considers the particular relevance of instruments that are not widely used in the evaluation of such programmes. It also raises the issue of whether evaluation can become so ritualised as to be valueless.

Case Study 3 is an account from the perspective of a non-participant observer (myself) of a once-a-week 'twilight' (after-school) course for ML teachers in Lothian secondary schools. The benefit to an evaluation of a non-participant observer's comments are one element in this study, but the main focus is on the use made of participant journals as input to other evaluation instruments.

Case Study 4 explores the effects of a tailor-made course for a group of German EFL teachers and trainers, as reflected in individual end-of-course action plans and subsequent reports on the implementation of these plans. The study suggests that action plans may not only prove a useful evaluation measure but also serve to prolong the period of reflection set in motion by the course, and therefore be a means of furthering programme objectives.

Case Studies 1-3 relate to continuing programmes; Case Study 4 is offered as a token of a programme type.

The salient features of the case studies are summarised below:

1 ML
* 2-hour TD seminar (in-house)
* content: use of overhead projector
* evaluative focus: pre-/post evaluation

2 EFL
* 3-week open-enrolment intensive summer course (60 hours)
* content: language practice/development; methodology
* 10-year study of evaluative procedures
* evaluative focus: rapid assessment form; effect of discontinuing this

3 ML
* 6-session (9-hour) course for state-school teachers
* content: mainly related to updating, seen from several perspectives
* evaluative focus: participant journals; non-participant observer

4 EFL
* 2-week (30-hour) tailor-made course for German EFL state-school teachers and trainers
* content: methodology (of teaching and teacher training); language and cultural updating
* evaluative focus: action plans

These four forms of INSET provision have been selected for their representativeness rather than their comprehensiveness. Other forms of provision, such as assessed, certificated programmes, are discussed in the context of the survey data and key informant interviews.

The final chapter, as is usual, attempts to weave together the various threads into a set of conclusions and recommendations.

1.8 SUMMARY

This chapter has offered a definition for programme evaluation, presented the case for programme evaluation and indicated the fragmentary nature of the literature in the area selected for study - evaluation of INSET programmes for language teachers. It has also indicated two ways in which the thesis seeks to make a contribution to knowledge and understanding: first, through a survey of current evaluation practice in relevant UK

institutions; and second, through the exploration and assessment of specific evaluation procedures.

The next chapter presents a framework for the discussion of issues and choices in programme evaluation. The perspective or angle of view adopted throughout this and subsequent chapters is that of the 'part-time' evaluator based in an institution and responsible for the evaluation of programmes with which s/he is directly involved, usually as a Course Director (CD), and/or with wider responsibilities for programme evaluation within the institution. In short, the perspective is my own.

CHAPTER 2

PLANNING AND IMPLEMENTING INSET EVALUATION

2.1 INTRODUCTION

Chapter 1 drew attention to the late emergence of a body of discipline-specific evaluation literature and the paucity of work on the evaluation of language teacher education. This chapter offers a more detailed critical review of the literature relating to the evaluation of INSET programmes. Although the main focus of the discussion is the evaluation within institutions of programmes for language teachers, reference is made where appropriate to the wider evaluation literature. The starting-point and framework for the whole chapter is a set of questions (see below) which highlight key concerns in any evaluation.

2.2 A FRAMEWORK FOR EVALUATION

2.2.1 Key questions

Reference was made in Appendix 1.1 to the various frameworks proposed as theoretical and operational models for programme evaluation. Frameworks have also been advanced in the field of second language education, but in this case they tend to amount to no more than a set of dimensions to be considered in planning and implementing evaluation in the form of a series of *Wh-* questions. The following summary of these questions draws on Rea (1983), McGinley (1986), Aubrey (1988), Mackay (1988, 1994a), Elley (1989), Hargreaves (1989), Abbey (1991), Nunan (1992), Rea-Dickins and Germaine (1992) and Weir and Roberts (1994).

WHY

... is the evaluation to be carried out ?

WHAT

... is to be evaluated ?
... criteria are to be used ?
... problems and constraints can be predicted ?

WHO

... will be involved in the evaluation, and what will be their respective responsibilities ?
... needs the information ?
... will see the report (s) ?

HOW

... will the evaluation be carried out (instruments, data sources) ?
... will communication be managed ?

WHEN

... will the evaluation take place ?

2.2.2 A decision hierarchy

Although the various *Wh-* questions tend to be presented in the literature as discrete items for consideration, they can for the most part be related in a sequenced manner, as indicated in the flowchart below:

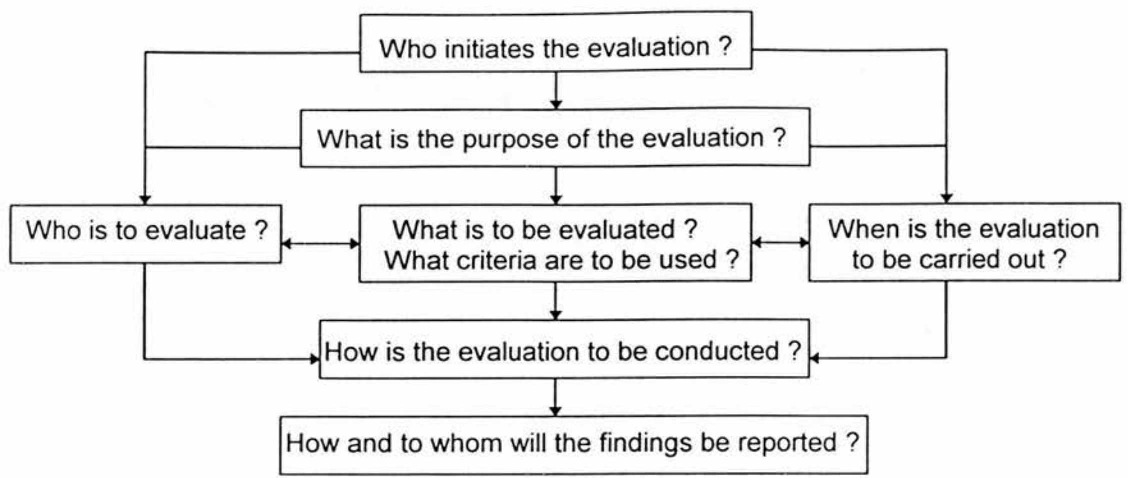


Figure 2.1: A decision hierarchy for evaluation

An illustration of how the framework would work in practice may be helpful. Let us suppose that the intention is to evaluate an INSET programme based within an institution.

Now although the *wish* to evaluate may manifest itself at the level of staff members directly involved with the programme, the *decision* to evaluate will rest with someone in a position of authority. In most public-sector UK institutions these days it will be an institutional requirement that all courses be evaluated as part of a routine procedure, yet the responsibility for ensuring that it actually happens still rests with one or more identifiable individuals.

The person or group responsible for initiating or overseeing the evaluation will also determine the purpose(s) that it is intended to serve. In broad terms, evaluation will be either accountability-oriented or development-oriented (see 2.3, below).The purpose (and those with powers for ultimate decision-making) will in turn affect the next layer of decisions. If the evaluation is to be purely developmental, then it is likely that much if not all of what goes on will be in the hands of programme personnel, although an outsider, possibly from within the institution, may also be called in as an adviser or neutral party. The issue of what is to be evaluated, the foci of evaluation, may be determined at the uppermost level, but it will certainly be influenced by the overall purpose and possibly by the concerns of the evaluator(s) (the two-way arrow suggests that the foci might also affect the decision as to who is to evaluate). The third decision at this level concerns the timing of evaluation. Again, the higher-level decisions will have an influence (managers have to allocate staff time and assess the value to the institution of time spent on evaluation; development-focused evaluation necessitates time spent on evaluation during a course). However, as the arrow from *What is to be evaluated ?* to *When ...?* indicates, the foci of an evaluation will also have implications for when it is carried out (compare the requirements of a focus on participant satisfaction with resources with a focus on programme impacts). Decisions concerning the distribution of time (and how much time is available) might equally have a constraining effect on what can be evaluated (hence the double-headed arrow). All the above lead to the question of method. If decisions concerning method rest with evaluators they will obviously wish to select procedures on the basis of their appropriateness to the questions asked (purpose, focus), but will normally be constrained by deadlines or time allocations; their own preferences and expertise will also be a deciding factor.

With the exception of Hargreaves (1989), none of the sources consulted offers a fully worked-out decision hierarchy of the kind outlined above, although some (most notably

Rea-Dickins 1991) give hints as to internal interrelationships. Hargreaves (*op. cit.*: 46), however, who uses rather different terms to represent some of the *Wh-* questions, shows how a framework which incorporates a similar sequence can be used to guide the evaluation of an INSET programme for a specific context (Appendix 2.1).

Each of the subsequent sections deals with one of the decision points in Figure 2.1:

WHY: Evaluation purposes (2.3)

WHAT: Evaluation criteria and foci (2.4)

WHO: The identity and characteristics of the evaluator (2.5)

WHEN: The timing of evaluation (2.6)

HOW: Choosing a method (2.7) - this incorporates:

- a framework for the comparison of evaluation methods
- constraints
- political and ethical considerations

HOW: Reporting evaluation (2.8)

The large issue of *How* ... ? in the sense of available methods and techniques for evaluation - a key chapter as far as the focus of the thesis is concerned - is taken up in Chapter 3.

2.3 EVALUATION PURPOSES

2.3.1 Evaluation for accountability and development

As the term implies, *evaluation for accountability* involves the presentation of evidence for 'outside interests' (Potts 1985: 25) in order to provide 'proof of the value of what has been achieved' (Ashworth 1985: 84) as manifest in the extent to which normative goals have been achieved (Murphy 1995). There may also be a concern to assess the relationship between the benefits and costs of a particular programme or form of provision (Alkin 1970; Rudduck, R. 1981; Dock, Duncan and Kotalawala 1988). This kind of evaluation will normally take place at a stage when decisions have to be reached concerning further investment (i.e. at an interim or review stage in the development of a programme) or when an assessment is needed of the value of a completed programme. In the latter case there may be implications for further investment in similar programmes.

The evidence presented in such evaluations will typically be 'hard' and take the form of observable products (e.g. attendance figures, test scores, numbers of people successfully completing the programme) since this kind of data can be relatively easily understood and related both to programme objectives of a quantified nature and budgetary considerations. Whether the evidence provides an adequate basis for decision-making concerning future programmes is another matter. Unless similarly hard data has been collected on prior occasions, it may be impossible to make comparative statements about gains; and if data on performance is gathered at too early a stage, it may be impossible to discern which effects are lasting or, indeed, which effects only emerge after some time. Moreover, without data on what actually happened, the processes involved, any attempt to establish causal relationships - a difficult enough endeavour in itself - will be purely speculative.

Data on effects and processes may be of no significance if programmes and evaluations are seen as being hermetic; it is important if one sees evaluation as a potential contribution to the sum of professional knowledge and, more specifically, as an input to the design of ongoing (Murphy 1995) or future programmes. In *evaluation for development*, the emphasis is less on passing judgements than on collecting information (Rea-Dickins and Germaine 1992), information that can be used for the benefit of the programme (*programme* development - see, e.g. McCabe 1980b) or the development of those involved as programme personnel or participants (*professional* development - see, e.g. McCabe 1980b; Rudduck, J. 1981; Murphy 1995). In essence, as Rea-Dickins and Germaine (*op.cit.*) note, evaluation for development examines the processes of teaching and learning (and the context in which these take place) in an attempt to identify those features that contribute to the success or otherwise of the programme. Among its potential positive outcomes can be the validation of existing practice and the dissemination through publication of good practice; a more commonly emphasised outcome (and specific purpose) is awareness-raising with regard to the need for change. This might be seen as a separate purpose (Rea-Dickins 1991; Rea-Dickins and Germaine 1992); however, as Rea-Dickins (1991) recognises, awareness-raising data feeds into the broader development process. Evaluation of the implementation of a curriculum innovation may also indicate the need for an in-service programme where this does not exist, and the desirable emphases of such a programme (Lawrence 1995).

As far as programme development is concerned - and this is particularly true of programmes in their initial stage - there is an obvious need for data of different kinds and from different sources. Test scores may be relevant, but they will only be one source of evidence to be considered, alongside observations and the comments of staff and participants (via questionnaires, interviews, discussions, and perhaps diaries), and much of this evidence will need to be collected while courses are in progress, either continuously or at agreed points. It is important, however, as Alderson (1992) points out, not to defer examination of the data until the course has been completed. Periodic analysis of the data during a course allows for changes to be made to data-gathering instruments or procedures as well as to the course itself, and such changes can not only affect the outcomes of that particular course but also accelerate the whole process of programme development.

Within institutions, there is likely to be a need for evaluation which serves both these purposes, accountability and development. Institutions are accountable for, among other things, the standards of the programmes they provide (a responsibility partly served in the case of award bearing programmes by an External Examiner or Moderator). They have a responsibility to the profession, programme participants and, if they are publicly funded, the public at large to meet their obligations as effectively *and efficiently* as possible. Without evaluation for accountability, resources may be under-exploited, inappropriately allocated or simply frittered away. Without evaluation for development, on the other hand, programmes and personnel may stagnate. Both evaluation purposes have particular significance in situations when a new programme is offered or a major innovation is introduced within an existing programme; but even in cases where programmes are well established periodic review which combines both accountability and development dimensions can be justified as an aspect of good management (Mackay 1988; Rea-Dickins and Germaine 1992; Wallace 1997).

2.3.2 Summative and formative evaluation

Various writers (e.g. Morris and Fitz-Gibbon 1978; Rea-Dickins 1991; Weir and Roberts 1994) relate the distinction between summative and formative evaluation (Scriven 1967) to the purposes discussed above. Summative evaluation, it is suggested, is a form of terminal stocktaking concerned with the provision of information to external bodies and is therefore a way of rendering an account to those who have a financial and/or policy interest.

Although some of the information on which summative judgements are made can be gathered during a programme the primary concern is with effects and information that is only available once a course is over. Negative findings may result in cuts being made or the programme being withdrawn (but see Mackay 1994a, reported below). The argument for summative evaluation is that (if results are positive) it is a way of justifying past and future spending; the argument against is that it may predispose the evaluation design towards the measurable (Weir and Roberts 1994). Formative evaluation, on the other hand, provides input to curriculum development (Rea 1983; Potts 1985) by providing information on e.g. materials, teachers or course organisation. If this information is made available during the course it can contribute to 'fine tuning' and therefore lead to greater effectiveness and short-term economies.

In practice, the differences between summative and formative evaluation are less clear-cut than suggested in this summary (Stake 1976). A summative evaluation may be carried out in respect of a single course, but if this is the first of several 'identical' courses within a programme, then the results of that evaluation may have a formative influence in very specific ways on subsequent courses (Bolam 1980). Similarly, if a course is part of an incremental series, the general lessons to be learned from one course may feed into the planning of the next. One way of characterising the relationship between summative and formative evaluation on the one hand and accountability-oriented and development-oriented evaluation on the other is shown below, in Figure 2.2:

	Accountability	Development
Formative		+
Summative	+	+

Figure 2.2: Contribution of summative and formative evaluation to evaluation purpose

There may be exceptions to the general rule implied in Fig. 2.2, as when an INSET programme incorporates formative evaluation (e.g. in the form of participant input to decisions on course delivery) and the effectiveness of this as a process is evaluated summatively; in general, however, formative evaluation contributes only to development, whereas summative evaluation can - if appropriate data is collected - be used for purposes of accountability or development. Although evaluation purpose is likely to affect the kinds of data collected, and the focus the methods used, the data itself is neutral. As Weir and

Roberts (1994: 5) put it, 'labelling evaluation data as formative or summative must relate to the purpose for which it has been collected. Where data are (*sic*) used to evaluate effectiveness against specified criteria, it is summative, and where it is used to influence change, it is formative'.

Rudduck, R. (1981) expresses some doubts about the extreme effects supposedly consequent on negative summative findings, citing the 'inertia' (1981: 6) and innate conservatism of the teaching profession. Mackay (1994a) similarly argues that in his experience of EFL/ESL programme evaluation the summative/formative distinction is unreal, since there is little evidence that negative findings result in a programme being terminated: 'Once ... in place, they generate their own impetus and their own political support' (1994a: 142). The alternative distinction that he proposes (*ibid.*) is between 'extrinsically motivated evaluation', which serves bureaucratic purposes (i.e. accountability) and 'intrinsically motivated' evaluation, which serves the purposes of programme personnel and learners (i.e. programme development). Extrinsically motivated evaluation is necessary, he suggests, not simply because the bureaucracy may require information that is not normally generated by intrinsically motivated evaluation but also because (Rea Dickins and Germaine notwithstanding) there is a dearth of evaluation models within the field of language teaching, and because there is little evidence that programme personnel carry out internal evaluation. Although it is not clear what the basis is for the latter assertion, Mackay's general conclusion carries a certain amount of weight: if programme personnel were seen to be carrying out internal evaluation, and if this produced the kind of information required by the bureaucracy (as it could) then there would be less need for extrinsically motivated evaluation other than that relating to broader policy issues.

2.4 EVALUATION CRITERIA AND FOCI

2.4.1 Introduction

Evaluation for accountability rests on the assumption that the value or otherwise of a programme can be assessed and that criteria are available for this purpose. The basis for accounting may be quantitative, as in cost-benefit analysis (Alkin 1970; Rudduck, R. 1981; Walker 1985; Dock *et al* 1988) or measures of efficiency, or it may be somewhat 'softer'

as in a concern for effectiveness. In evaluation for development, the criteria may also be course-specific. The importance of this level of the decision-making hierarchy cannot be overestimated. As Hodgson and Whalley (1985) point out, evaluation criteria determine the questions the evaluator sets out to answer and hence the data to be collected and the use that will be made of that data. This section begins by discussing some of the problems connected with criteria for evaluation; it then looks at the related issue of evaluation foci and how these might be determined.

2.4.2 The quest for suitable criteria

It is generally agreed that criteria are necessary for evaluation. Hargreaves (1989) acknowledges that some studies might be purely exploratory, but dismisses these as evaluation 'since evaluation necessarily entails making judgements about past and/or future action - and this implies the existence of criteria or expectations against which such judgements are made' (39). Earlier writings on INSET evaluation voice concern, however, with respect to the nature of the criteria to be used, e.g. in the assessment of *effectiveness* (Spelling 1981).

If the ultimate goal of INSET is to benefit the pupil in the school (rather than the participant - and this is a moot point) then the ideal way of measuring the effectiveness of INSET programmes would seem to be by pupil-based enquiry. This is a little idealistic (Cooper 1983). Bolam expresses the general view: 'However desirable it may be to establish the impact of INSET on pupil behaviour or achievement, this is unlikely to be technically possible in most situations' (1980: 45). The technical factors Bolam may have in mind include the difficulty of controlling for contextual variables in order to establish direct cause and effect relationships at either of two levels: that of the impact of INSET on the teacher or that of the impact of the teacher on pupils; what makes this particularly difficult, perhaps, is that when teachers leave an INSET programme they find themselves in different contexts - contexts which will be more or less receptive to innovation (Ross, McNamara and Whittaker 1977; Dock *et al* 1988). Moreover, and perhaps as a consequence, teachers may not teach in the way they were expected to following the programme (Elley 1989). Ross *et al* (*op.cit.*) and Baker (1980) point out that even if it were possible to attribute change to the effect of the INSET programme it would be difficult to isolate those features of that programme which had been particularly influential.

The difficulty is not merely a technical one. Baker sees a problem in the absence of an adequately developed construct of effective teaching and of research tools for investigating this; to judge from later writers (e.g. Peck 1988, Long 1990), no significant advances have been made in this respect.

Joyce and Showers (1980, cited in Aubrey 1988: 143) distinguish four levels of impact for INSET:

1. awareness
2. acquisition of concepts and organised knowledge
3. learning of principles and skills
4. application of these to classroom teaching.

They make the incidental point that only when the fourth level of impact is achieved will it be worth looking for pupil effects. Whatever one may feel about the feasibility of looking for indirect, pupil effects (see above) the value of such a hierarchically ordered list is twofold: it allows for the possibility that INSET may have different sorts of effect (on different individuals, possibly) and for the assessment of the programme's effectiveness at any of these levels.

One view put forward is that criteria should be based on programme objectives (Hodgson and Whalley 1985). Hodgson and Whalley argue that evaluation is only valid if (1) the criteria are relevant and appropriate and that (2) the list of criteria is complete. One way of generating a list of criteria would be by brainstorming, an approach which they reject; the other, which meets with their approval, is to derive criteria from judgements about the course and its explicit aims and objectives. Hargreaves (1989), Alderson (1992) and Weir and Roberts (1994) all express reservations regarding the suggestion that the explicit aims and objectives of a course should form the basis for decisions concerning the criteria for its evaluation, pointing out that programmes have their own dynamism and objectives may be implicitly modified over time. There are other weaknesses of objectives-based criteria. Effects may not be observable within the timescale specified for evaluation; effects which are neither looked for nor apparent to programme personnel may be missed; and, as noted in Appendix 1.1, the assumption that the objectives of a programme are necessarily valid, consistent and appropriate may itself be suspect, which means that the criteria derived from these objectives will also be suspect. An explicit or underlying concern for participant satisfaction is the most obvious exception to this.

The second term used in considerations of criteria for programme evaluation is *efficiency* (Murphy 1985; Dock *et al* 1988; Abbey 1991). In evaluating efficiency, the main considerations will be costs (time, money) in relation to benefits[10]. The efficiency of an award bearing programme might thus be determined by calculating the pass rate as a proportion of those enrolled; although one might also wish to take into account such factors as dropout rate and the time required for successful candidates to complete the course. Successful course completion may also be perceived as an aspect of programme effectiveness. Dock *et al.* (*op.cit.*) report on two INSET programmes for non-graduate, untrained teachers in Sri Lanka the primary purpose of which was to reduce the backlog of untrained teachers. Given this purpose, it is not surprising that the *effectiveness* of the three-year distance-learning programmes is discussed with reference not only to the impact of the programme on classroom teaching (as reflected in changes in teaching methods, changes in teachers' attitudes, increase in teachers' knowledge, and student achievement) but also to graduate output.

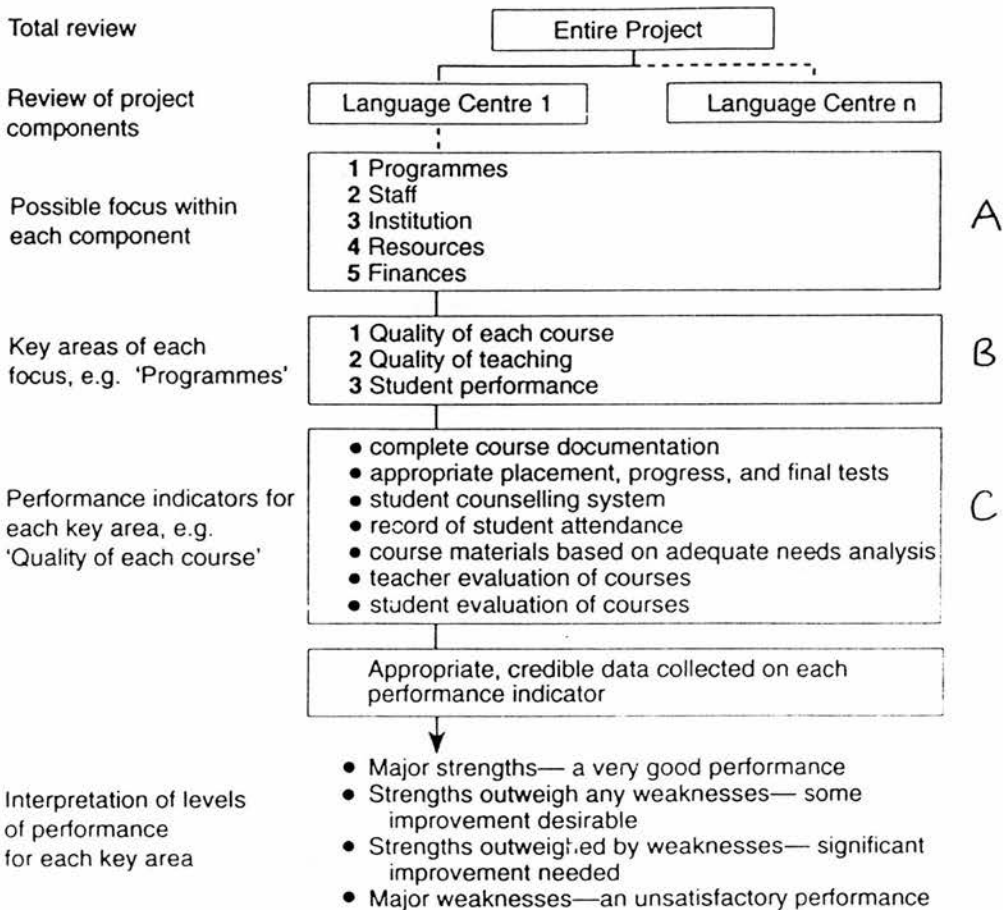
2.4.3 Criteria in project evaluation

The question of criteria has been of particular relevance in relation to project evaluation and since educational projects often include a teacher education element, some reference to the literature of project evaluation may at this point be relevant.

Project frameworks are designed in such a way that the project can be evaluated at various stages and according to different types of criteria (see, e.g. Appendix 2.2). In the case of externally-funded accountability-oriented evaluation, it would be normal for the funding body to establish the criteria, perhaps in negotiation with the evaluator. For ODA projects, these are expressed in terms of 'indicators of achievement' of the objectives specified (as in Appendix 2.2).

The practical value of such a framework is, however, questioned by Chambers and Erith (1990), who raise the issue of level of effects. The benefits that accrue from projects, they say, are of two kinds: products, which are quantifiable, and effects, which may be much more difficult to assess because less easily observable.

The scheme put forward by Mackay (1994a) for the evaluation of language centres in Indonesia appears to avoid this problem (see overleaf). Mackay is here concerned with the evaluation of language programmes, but the approach seems equally applicable to any kind of programme:



Mackay 1994a: 147 (*letters added*)

The upper part of the diagram (Boxes A-B) sets out in hierarchical fashion a range of possible evaluation foci relating to the context. Criteria for the evaluation of 'programmes', one of the key areas identified, are set out as performance indicators. Performance on each of these indicators is then assessed by programme personnel using the descriptors of levels of performance at the bottom of the diagram.

In operational terms there may be certain problems with this scheme. At the level of 'key areas', the assumption is that data can be collected which will permit 'measurement' of 'adequacy' and 'effectiveness' (148) using the descriptors of levels of performance. Yet decisions about how to grade specific features will clearly be subjective unless there is further discussion as to what constitutes adequacy or effectiveness in relation to each feature. Nevertheless, Mackay's scheme has much to recommend it. It is systematic, context-sensitive and responsive to the perceptions of programme personnel, who have generated the categories in Boxes A-C.

The relevance of project evaluation to institutional programme evaluation lies not in the criteria used but in the hard-headed demand for evidence that the programme has worked. The fact that this may lead to 'bean counting' or the collection of evidence that is relatively superficial is not in itself an indictment of the system but a product of the short-term nature of projects. Applied to the longitudinal study of programmes, a similar approach should yield much more interesting results.

2.4.4 Foci of evaluation

In relation to the focus of evaluation two general questions arise: the first is the perhaps deceptively obvious 'What is to be evaluated?', the second 'How are decisions regarding focus to be reached ?'.

2.4.4.1 Levels of foci

The first question can, in fact, be answered at different levels of specificity, as a number of writers have recognised.

Fox (1980) categorises INSET purposes as (1) improvement of school practice (2) professional development (3) implementation of social policy, and contends that a concern for the ultimate purpose of the education and/or training should determine the focus of evaluation. Thus, if the purpose is the improvement of school practice, then the focus of evaluation should be not the programme itself, as is normally the case, but the relationship between the programme and school practice - and specifically the extent to which and ways in which the programme appears to have contributed to changes in practice. One relevant input to such an evaluation, he suggests, would be participants' views, not on the course itself but on the effect the course they have followed has had on their practices. This framework for decisions concerning foci is clearly specific to INSET; other approaches are of more general relevance.

Thus Bolam (1980), cites Stufflebeam *et al.* (1971), who propose the following set of categories:

1. context (assessment of needs and the match between these and the goals of current systems)
2. input (how to utilise resources to meet programme goals)
3. process (who does what to whom, when, where and how)
4. product (programme effects).

Alderson (1992), writing about the evaluation of language programmes, offers this categorisation of possible evaluation foci:

1. outcomes (e.g. language proficiency, attitudes, behaviour)
2. impact on context
3. features of the programme and processes
4. teacher training implications and activities
5. resource implications
6. relationship between costs and benefits.

Weir and Roberts (1994: 18) reproduce a set of focus categories from Sanders (1992) which can be summarised as:

1. programme needs assessment (related to the setting of programme goals)
2. individual needs assessment
3. resource allotment
4. processes or strategies for providing services to learners [this item is further specified]
5. outcomes of instruction.

These categories are, as Weir and Roberts note, largely developmentally-oriented. For example, several of the sub-categories within item 4 are expressed in terms of 'insights' (e.g. *b. classroom processes: to provide insights about the extent to which educational programmes are being implemented*). The substitution of 'evidence' for 'insights', Weir and Roberts point out, would have a significant effect on the data required and the presentation of findings.

The similarities between the three lists are quite striking. For instance, programme *effects*, or outcomes, appear in all three lists, as do *process* features and *resource* considerations. For the rest, Alderson's item 2, 'impact on context', might be seen as a particular kind of outcome, but also implies some prior assessment of the *context* (see the first items in the lists of Sanders and of Stufflebeam *et al.*). His item 6, 'relationship between costs and benefits', is tied to consideration of outcomes, but is an important supplementary question. His item 4, although specific as it stands to the evaluation of language programmes rather than INSET programmes, draws attention to the fact that evaluation may reveal the need for certain kinds of continuing staff development. Alderson's list would therefore seem to be the most comprehensive of the three, though it might give the impression of greater coherence if reordered along the lines of the other two lists.

Within the kinds of category identified above, the range of possibilities is immense. For example, Tribble (1993) gives a brief description of INSET evaluation in the Baltic States which looked at administrative and support mechanisms (recruiting procedures, facilities and resources, institutional support) as well as the training programme itself, the course tutor, changes in participants and changes in the pupils taught by participants. He notes that this study was partly intended to identify other areas for 'more focused evaluations' (1993: 2). The fact is that 'almost anything can be evaluated ... from materials to methods,

from costs to values, from stress loading to creche facilities' (Woodward 1991: 220). For this reason, decisions taken at the level of purpose and of categories such as those discussed above may be more principled and coherent than those at a lower level of specificity.

Although the implication in previous paragraphs has been that it is the evaluator who decides what to evaluate, ideally in consultation with others, this will in practice depend on such factors as the overall purpose of the evaluation and the role defined for the evaluator in relation to that purpose. As indicated by the decision hierarchy in 2.2.2, one decision leads to another. The significance of decisions concerning evaluation foci becomes clear when one realises that the selection of foci has direct consequences for what can be learned (Woodward 1991). One conceivable outcome of evaluations oriented specifically towards accountability may be that little information is forthcoming of direct value for programme development and *vice-versa* (Mackay 1994a).

We now consider how decisions on foci can be reached.

2.4.4.2 Deciding on foci

Rudduck, J. (1981) lists four questions that the would-be evaluator should consider before deciding whether to evaluate a short course:

1. Are there any specific problems, uncertainties or novelties that the evaluation will be expected to illuminate ?

2. Will the data gathered genuinely contribute to the improvement of future practice ?

3. Will the conduct of the evaluation be likely to illuminate the critical reflection of course members, or be of other benefit to them ?

4. Could the data gathered usefully be organized and presented as a formal document for a wider audience ?

(Rudduck, J. 1981: 50-51)

If the decision is taken to go ahead with the evaluation, then Questions 1 and 2 will, Rudduck observes, help to sharpen the foci.

Programmes evolve over time and it is logical that the foci of evaluation should keep pace with these changes. In determining priorities for developmental evaluation of courses

within continuing programmes the evaluator can begin by assessing what appears to be known and can be taken for granted, i.e. the features of the course which have previously been consistently evaluated positively, and isolate those features which have been subject to revision or merit further investigation, perhaps using other evaluation procedures. This agenda might then be modified by discussion within the course team and if appropriate with participants, as Reid (1994) suggests.

Cronbach (1982: xi) proposes a four-factor approach to the prioritisation of foci within long-term programmes in which the first two factors are reminiscent of Rudduck's (*op.cit.*) Questions 1 and 2:

1. *prior uncertainty* (Is there any real doubt ?)
2. *information yield* (How much will we learn ? How much will remain uncertain ?)
3. *cost* of enquiry (Time ? Money ?)
4. *leverage* (Is the information capable of influencing operating decisions ?).

Cronbach's view is that factors 1 and 4 should take precedence. Although it is not difficult to justify the inclusion of these factors, their proposed pre-eminence is at least arguable. It would be a rare evaluation in which cost (item 3) was not a primary factor and in practice, the relationship between items 2-4 is quite complex. One would have to consider, for instance, whether the likely information yield (2) would justify the allocation of resources (3), and what it would cost (3) to provide information with which leverage could be exerted (4). In other words, there is an issue of sequencing and the effect of a negative response at any point in the sequence. Figure 2.3, below, is an attempt to resolve some of these problems:

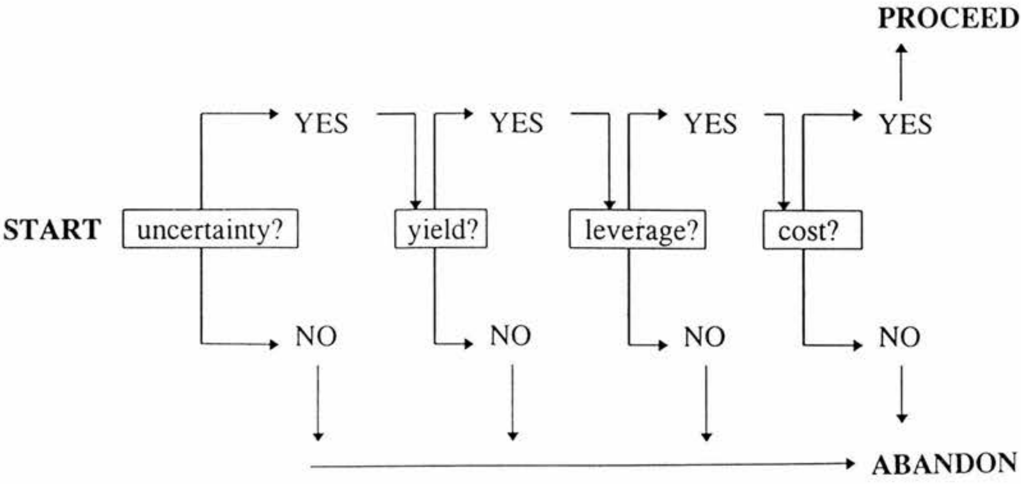


Figure 2.3: Prioritising evaluation

A further loop in the flowchart is available. If the cost of the initial plan is deemed unacceptable the plan may be trimmed down and questions of yield, leverage and cost reconsidered.

One aspect of the potential for action consequent on evaluation (leverage) is also highlighted by Mackay (1994), who emphasises the importance of restricting evaluative activity to those features of a programme which are within the control of programme personnel. An obvious consequence of an evaluation which is not focussed in this sense, he claims, is that programme personnel become frustrated. This possibility exists, of course, but it is also possible that an outside evaluator (influenced or not by programme personnel) might wish deliberately to draw attention to constraints which prevent the programme from achieving its intended effects, and this could in time lead to the removal of these constraints.

2.5 THE IDENTITY AND ROLE OF THE EVALUATOR

2.5.1 Introduction

In theory and perhaps ideally, evaluators would be selected on the basis that they possess the qualities and skills necessary for the task in hand (i.e. the movement within the decision hierarchy would be from *What is to be evaluated ?* to *Who is to evaluate?*). Within institutions, where responsibility for evaluation tends to be job-related, this is less likely to be the case. Either the CD, say, by virtue of his/her position, takes decisions concerning what and how to evaluate or s/he simply applies institutional procedures (e.g. (standard questionnaire, report form).

The emphasis of this section is less on the skills needed for evaluation than on the identity of those involved in institution-based evaluation. Should this be one person or more than one ? What should the relation be between this person or these people, the programme to be evaluated and the individual or group initiating the evaluation ? The related question of evaluator role is also discussed, however, since this may affect decisions concerning the choice of evaluators and the way in which responsibilities are specified.

2.5.2 Participation

The desirability of all with a vested interest in a programme participating in its evaluation, and constraints on the implementation of this principle, have been discussed by *inter alia* Fox (1980), McCabe (1980b, 1987), Thorpe (1988), Hargreaves (1989), Alderson and Scott (1992) and Mackay (1994b). All are in agreement with Henderson (1978), who asserts that evaluation is not something done *to* but *with*.

While generally sympathetic to the principle of involvement not only of INSET programme participants (as adults and fellow educators) but also of administrators and possibly pupils as the ultimate beneficiaries, Fox (*op.cit.*) points out that the availability and characteristics of participants may limit their participation; in some cases, moreover, their willingness to cooperate fully may also be in doubt. Specific problems in this respect would be lack of relevant previous experience of evaluation or negative experiences of evaluation. Bolam (1980: 43) recalls House (1972), who was obviously thinking of 'inspectorial' evaluation: 'those on the receiving end ... have little to gain and a lot to lose'. Where expectations (of staff 'on the receiving end') are negative, avoidance and threat-reduction strategies can be expected or attempts to co-opt the evaluator into the programme and its value systems. Sharp (1990), in the context of a paper arguing for staff-student cooperation in evaluation, also voices doubts which might be relevant to INSET evaluation in certain contexts. These include the fairly obvious possibilities that responses might be untrue (students give what they assume to be the desired answers) or incomplete (they may fear an adverse effect on their grades) and therefore misleading, but also a risk in terms of tutor 'face' (to ask for students' comments may give the impression that tutors do not know what they are doing) or reputation (students may vent personal antagonism).

In projects or long-term teacher education programmes which see teachers' capacity for evaluation and self-evaluation as the key to professional development these difficulties might be addressed through programme content and process; on short courses, participation in a full sense may be ideologically sound but practically unrealistic. Fox's general conclusion, that the nature and extent of participation should ultimately be determined in relation to evaluation purpose and context, seems an appropriate compromise.

Later commentators tend to be more enthusiastically committed to the notion of participation, especially in relation to project evaluation (e.g. Alderson and Scott 1992). In this, they may reflect the politically correct stance of the late 1980s, a stance aptly summed up in the following somewhat tongue-in-cheek quotation from McCabe (1987):

> Everyone's view is important. Evaluation is a continuum embracing everyone in the learning business, teachers, pupils - and even evaluation consultants. Evaluation, according to this ideal view, involves everyone in learning more deeply than before. Its harnessing makes everyone's view valuable, and it makes alienation less likely, in the sense that students or teachers who feel that they can influence what is being done, who know that their views are being heard are more likely to feel responsible for and committed to their education.
>
> (McCabe 1987: 1-2)

The notion that 'everyone's view is important' is contentious not because it is 'ideal' but because it takes no account of the relevance of the views of specific groups or individuals to the purpose of an evaluation. Participation is an important principle, for the reasons given by McCabe and in subsequent subsections, but it is not a principle to be applied blindly.

2.5.3 Insider or outsider evaluation

In practice, what tends to determine, in broad terms, who will carry out an evaluation is evaluation purpose. Most evaluation for development is carried out by those with direct involvement in the programme. These are the people who have a special interest in seeing how far the programme has achieved its aims and what the implications are for the detailed planning of future courses. Knowing the institution and the course, they are in a position to work rapidly and at minimum expense (Beretta 1990a, citing Patton 1986). However, they are not necessarily the most appropriate people to collect and process the kinds of evaluation data that would permit informed decision-making. Some reasons for this are set out by Rudduck, writing in the context of courses organised by LEAs:

> The fact that a course organizer has set up a course gives him or her authority and establishes or enforces a status difference *vis-à-vis* the course members. While it may still be possible for participants to feel that they can comment openly in writing or complete an assessment form, they may not feel easy about offering spoken comments, either individually or in a group situation. Moreover, the course organizer, even if there were no problem in eliciting personal comment, may have insufficient time to explore the detailed components of the course members' criticism. For these reasons, we have tended to seek outside commentators who will sit in on a course.
>
> (Rudduck, J. 1981: 70)

There may be no obligation to involve someone external to the course team in the evaluation of non-award bearing courses and yet, as Rudduck suggests, there are still various advantages in this. Other reasons for involving an 'outside commentator' are discussed below, although care needs to be taken in relating these comments to the particular context to which they refer. (Note that the degree of 'externality' also differs.)

Rowntree (1985: 245) sees a value in the 'critical friend', who is not really external since they work within the institution and presumably share in its educational ethos. However, because they are 'with the group yet not of it' (*ibid.*) they can act as an objective referee and help to ensure that issues are properly talked through. Mackay (1988), on the other hand, is clearly thinking of a professional evaluator when he argues for the involvement of an external evaluator on the grounds that such a person possesses skills that may be lacking in the course team; may introduce a new perspective on specific issues; and may be less constrained by the interpersonal and political pressures which not infrequently build up inside a programme instruction team. Writing about fairly large-scale summative evaluations, Elley makes the point that evaluators should be independent 'so that no vested interests are involved *or perceived to be involved'* (1989: 270-271, emphasis added); if no suitably qualified independent person is available, he adds, the evaluation should at least be planned by a committee, with data collection and analysis being carried out by neutral researchers. For Alderson (1992) the decision as to whether or not to call in an external evaluator may be affected by internal sensitivities: some situations may be so sensitive that an outsider cannot be involved; in other cases, insiders may be so divided that the presence of an outsider is essential. The impartiality of an outsider is, he cautions, no guarantee of his or her objectivity; the fact is that no evaluation can yield the Truth in any absolute sense.

One argument for the use of an external evaluator is, then, that they bring an independent perspective and neutrality to the scrutiny of what is or has been happening. An argument for the use of a *professional* evaluator is that the latter also has special skills which will not only be of intrinsic value but, employed in the context of participatory evaluation, may also benefit those with whom the evaluator works (Davis 1980; Elliott 1980; McCabe 1980b; Thorpe 1988). In an institutional context, however, the costs of the latter will be an obvious deterrent.

2.5.4 Subject specialist or generalist

Where an individual rather than an evaluation team is given the responsibility for conducting an evaluation (a decision which rests on the scale of the evaluation and/or the skills required), there is an issue as to whether this should be a specialist in evaluation or a subject specialist, if the choice is as stark as this. McCabe (1980c) considers that an academic background in psychology or sociology is desirable, but warns against the effects if either orientation is too extreme. Beretta (1990a), who like McCabe is thinking of professional evaluators, also comes down on the side of the generalist, albeit reluctantly. On the basis that there are at present too few people within second language education who are trained evaluators, he concludes that it would be preferable to use professional evaluators since the latter can relatively quickly grasp the basic concepts of a new discipline. Leaving aside the question of cost, one major snag is that this kind of outsider would not be capable of evaluating certain specific features of a programme, such as the appropriateness or quality of the content (Rowntree 1985), and might in general lack credibility in the eyes of those running the programme or those responsible for acting on an evaluation report. The ideal situation, Beretta argues, would be if more language teaching professionals were also trained evaluators, a point taken up in a recent colloquium on the evaluation of projects (reported in British Council 1995). The implication of all this is, of course, that the kind of institution-internal evaluation currently being carried out is amateurish, or at least uninformed, a point to which we return in Chapter 4.

Marked differences exist, of course, between small-scale institution-based programmes and the typical overseas project in language education or teacher education. There are a number of predictable reasons (cost, market-sensitivity, participant reaction) why an

institution wishing to evaluate small-scale *non-award bearing* programmes would not consider calling in outsiders, whether subject specialists or generalists. The benefits of involving a 'critical friend' from within the institution might, however, seem quite attractive. This brings us to the way in which such a person might be used and their desirable characteristics.

2.5.5 The role of the external evaluator

The evaluator has been compared to, among other things, a 'tradesman' (McCabe 1980c), a 'craftsman' (Fox 1980), an 'inspector' or 'management consultant' (McCabe 1987), and a 'watchdog ' or 'hired gun' (Borg and Gall 1989: 745)); and the responsibilities of evaluators to the academic community (evaluator as researcher) and to the policy-shaping community have been discussed by Beretta (1990b). The central issue underlying such discussions is, of course, that of the independence of evaluators and the nature of their role and responsibilities.

McCabe's (1980c) evaluator is a tradesman in the sense that s/he is a professional, working within someone else's structures and paid for their time, materials and expertise. The way such an evaluator satisfies an employer is by providing 'credible, understandable and useful answers' (Mackay 1988: 41) to the questions of the (principal) stakeholder(s). One problem with this position is that it may lead the evaluator to disregard the interests of those stakeholders, including programme personnel, who were not responsible for commissioning the evaluation. Another problem, and this is brought out clearly by Beretta (1990b) is that the stakeholder's initial questions may prove to be unanswerable or answerable in only the vaguest terms, and it therefore falls to the evaluator to formulate appropriate questions in consultation with all stakeholders. The real issue, then, is who determines the questions to be asked and the consensus view - admittedly that of writers who are themselves evaluators - seems to be that this is the responsibility of the evaluator, as are decisions concerning the design of the evaluation, the specific techniques to be used, and the product, although it is recognised that it would be sensible to negotiate these with all who are likely to be affected. Where an external evaluator has been imposed by the bureaucracy rather than invited in by the course organiser, communication takes on particular importance, since the evaluation may well be perceived as threatening.

Saville and Andrews (in Rudduck, J. 1981) describe some of the options open to an external evaluator who is to observe a course:

i to act as an observer, offering a personal interpretation of events and experiences;

ii to take a semi-participant role where the observer comments partly on the basis of his or her own experience of being a member of the course but is also free to elicit and communicate the views of participants;

iii to act entirely as a mouthpiece for the participants, representing their views of the course rather than his or her own;

iv to act as a resource and consultant to the participants and the course organizers, offering informal commentary as the occasion allows.

(Rudduck, J. 1981: 71)

It is important, they stress, that the role of the evaluator is clear to participants, in particular to allay their natural suspicion that they are themselves the object of evaluation:

> the role, brief and identity of the evaluator have to be made very clear to the participants so that their behaviour on the course is not modulated by anxiety or uncertainty about the nature of the evaluation task.

(Rudduck, J. 1981: 72)

This was brought home to them, Saville and Andrews explain, by a negative experience with an evaluator who failed to maintain a consistent role and also lacked the normal 'paraphernalia' (notebook, tape recorder) of the evaluator. They contrast this with an evaluator who formally explained his own role to participants and how he intended to fulfil that, and was armed with the evaluator's 'insignia' (72-73).

As this subsection has indicated, the question of who should evaluate is not as straightforward as it may seem to be to those in the field - managers wanting to make financial decisions on the basis of achievements or programme personnel seeking information for programme development purposes. There are, it is clear, advantages in the involvement (paid or otherwise) of one or more people external to the course team in both accountability-oriented and developmentally-oriented evaluations; equally, there is a strong case to be made for the involvement of programme personnel (and others, perhaps) not

only in evaluation for development but also - and not simply as informants - in evaluation for accountability. Few would disagree with Hargreaves' view that 'the closer the collaboration between internal and external agents, the more thorough and effective the evaluation is likely to be' (Hargreaves 1989: 42). What this implies is that those initiating evaluations need to give careful thought to what an evaluation can provide and how the answer to that question might affect the choice of evaluators.

2.6 THE TIMING OF EVALUATION

2.6.1 Evaluation in planning for INSET

As Hargreaves (1989) and others have pointed out, evaluation is often seen as the last stage in a linear process:

DESIGN → IMPLEMENTATION → EVALUATION

At least two consequences follow from this sequential approach. One is that evaluation may not take place at all because the priority at the end of one course is to prepare for the next rather than spend time thinking about the last. The second consequence is that if evaluation begins only when implementation is at an end relevant data (baseline, process) may no longer be available.

A more appropriate approach, Hargreaves contends, is one that conceives of evaluation as an integral part of the design process and incorporates an evaluation framework from the beginning. This might have at least three effects, two of which are related. It would permit the necessary data to be collected, thereby allowing a course to be modified while in progress (Hargreaves, *op.cit.*). It might also encourage programme personnel to adopt a more reflective attitude to their teaching and other aspects of the course than might otherwise have been the case, and this might in turn generate data (e.g. in the form of records or diaries) that might not have been anticipated.

Hargreaves' proposal seems perfectly sensible and yet to judge from the comments of such well-travelled ELT practitioners as Potter (1994) and Murphy (1994) there appears to be little evidence that his views have been widely adopted, even in the context of projects. One reason may be that although lipservice is paid to the value of evaluation, the conditions necessary for its effective functioning have gone unrecognised or been ignored.

There may be some circularity in this, however. It is possible that evaluation has failed to prove its worth precisely because it was ill-conceived (e.g. the wrong questions were asked or the requisite data was not available).

Even when evaluation is built into a programme or project in the way that Hargreaves envisages, and even when it has a post-programme dimension, there are still problems associated with the time-scale for the measurement of effects. As Alderson (1992) acknowledges, since we are unsure what to expect, we do not know how long to evaluate for. The consequence is that we do not know how short term the observed effects of INSET may be; nor, on the other hand, do we know how long it takes for certain types of effect to appear. If open-ended evaluation were an option, other questions could also be explored, such as the relationship between attitudinal and behavioural change and whether the adoption of ideas in time leads to their adaptation. One problem with the concept of open-ended evaluation is that over some years programmes change almost of their own volition, and 'an evaluation that takes too long is likely to be overtaken by events; and therefore to become irrelevant' (Alderson, *op.cit.*: 294). Valid though Alderson's point may be in relation to evaluation for the purposes of development it takes no cognizance of the potential *educational* value of evaluation (see 2.3). It seems quite likely that what happens after courses is more generalisable than what happens during courses and that even a small number of contextually differentiated long-term studies focussing on implementation would have immeasurable illuminative value for INSET programme design in general.

Thus far, the impression may have been given that it is evaluation purpose that determines the timing of evaluation activity. Accountability-oriented evaluation thus takes place at the end of a course because this type of evaluation is concerned primarily with what has been achieved, whereas developmentally-oriented evaluation will normally be continuous, involving both staged formal evaluation and informal monitoring. Leaving aside purpose, at least two other factors should play a role in decisions as to when evaluative soundings are taken. The more obvious of these relates to the resources available - the extent to which it is feasible (irrespective of whether it is deemed to be desirable) to carry out baseline data collection, for instance, or follow-up studies. The second factor is the criteria by which the effectiveness of the course will be judged. If these include some reference to behavioural effects (e.g. changed teaching practices) then it will be important that the

evaluation design contain a post-course stage in the form of observation or reports, for instance; but equally it will be important that some measure be taken of pre-course behaviour so that judgements can be made on the basis of observational evidence rather than assumptions or retrospective self-report. For the time being, it seems that the best that evaluators can do is to attempt to set up evaluations in which data is collected either continuously or during 'meaningful episodes' (Weir and Roberts 1994: 17) from the beginning of the course or, if relevant, before it begins, until an agreed point after the course is over. 'Evaluation is done to avoid being wise after the event', as Murphy (1995: 13) puts it.

The remainder of this section discusses issues relating to the following evaluation stages:
ñ pre-course (preliminary) evaluation (2.6.2)
ñ formal in-course monitoring (2.6.3)
ñ end-of-course evaluation (2.6.4)
ñ post-course evaluation (2.6.5)
Particular attention is paid to pre-course and post-course evaluation since, as will become clear from the discussion below, these are crucial stages which may on occasion be somewhat neglected. The section ends with a brief consideration of informal monitoring (2.6.6).

2.6.2 Pre-course evaluation

The assessment of, for example, participants' needs, expectations, knowledge and skills prior to or at the beginning of an INSET course can serve two purposes, excluding that of selection: such information can lend specificity to course design, making a good match more likely between course provision and participant needs and wants; it can also serve as a yardstick or baseline against which change can be assessed at predetermined stages during a course, when it might be thought of as 'milestone baseline data' (British Council 1995: 100), or following a course. The importance of baseline data has not always been recognised, yet without some reliable indicator of the entering behaviour (prior knowledge, skills, attitudes) of participants, there can be no valid basis for claims about the effects of courses (Hargreaves 1989; Alderson 1992; Rea-Dickins and Germaine 1992).

Thorpe (1988) offers a rather different perspective on the nature of baseline data and how and when it might be used. For her, baseline data is collected during a programme (i.e. from a series of successive courses) in order to establish the stable features of the programme; subsequent analysis of this data reveals which features merit more focussed evaluation.

Evaluation - as the reference to Thorpe illustrates - is not only concerned with the effect of courses on individuals; it is also concerned with the effects of changes that are made to programmes in response to evaluative activity. Weir and Roberts (1994) make the point that if a course is part of a series (i.e. a programme) and changes are continuously being made in the light of successive evaluations, there is a case for staggered baseline studies to be conducted to assess the impact of these changes.

Indirect effects, in the form of changes in the learners taught by INSET participants, may also be a focus for evaluation, although this is generally agreed to be difficult (see 2.7.3.1). While fully aware of the difficulties associated with pre- and post-tests, Weir and Roberts (*op. cit.*) undertook an evaluation of an INSET programme in Nepal in which pre- and post-test data was collected to assess the effect on learners. Early in the study, they realised that they would also need data on the use participants made of the training they had received:

> without these data it would not have been possible to interpret the impact of the project ... if there had been no significant difference in performance between pupils in trained and untrained teachers' classes, we would not have known why this was so. The teachers may not have been putting into practice essential elements of the training and we would not have been aware of this.
>
> (Weir and Roberts 1994: 16)

Although the focus of the quotation is not the pupils, it will be obvious that a two-tier and two-stage comparison was actually called for: between both teachers and pupils before and after the INSET programme. In respect of the teachers, one obvious difficulty in collecting baseline data which relates to teaching practices is that it needs to represent adequately the population from which participants will be drawn and to be a fair reflection of their normal practice. Although this could include questionnaire and interview data, it would also entail

observing not just an appropriate sample of potential participants, and even non-participants, but multiple observations (before and after training) to ensure that judgements are as reliable as possible. In situations where the target population is large and dispersed, resource constraints are likely to make this kind of evaluation design totally unrealistic.

Other forms of pre- and post-comparison may be more feasible. For instance, a consideration of participant performance on an award bearing course may be related to the personal information (e.g. test scores, qualifications, experience) collected prior to the course for the purpose of selection. This may lead to the conclusion that certain types of data accurately predict performance or, conversely, that performance cannot be predicted from the kinds of data that is normally collected - and that other selection procedures need to be considered. (This point is further developed in Chapter 3.) Of interest to sponsors is the extent to which participants are subsequently able to implement what they have learned, and where post-course feedback indicates little or no transfer of training, a case could be made for providing institutions making pre-course contact with institutions that are releasing staff with a view to negotiating the use that will subsequently be made of these staff (Marsh 1987). Such pre-course negotiations or feasibility studies might in this way assess institutional as well as individual receptivity to change. The importance of understanding context and taking account of attitudes to change has been recognised in respect of non-specific INSET (Hopkins 1989) and TESOL project design (see e.g. Holliday 1992 and Ainscough 1994), but it also has relevance for UK institutions offering INSET programmes for teachers from overseas and those funding these teachers.

2.6.3 Formal in-course monitoring

The most obvious reason for conducting evaluation while a course is in progress is so that modifications can be made to it for the benefit of those currently involved. Judging by the literature, the main source of information is participant feedback, oral and written. This will typically come from questionnaires (see, e.g. Henderson 1978, 1980, 1981; Rudduck, J. 1981; Parker and Graham 1995; Ward *et al.* 1995) or from group discussions (see, e.g. Rudduck, J. 1981; Rowntree 1985; Morrow and Schocker 1993; Wallace 1997); participant diaries may also be used for this purpose (see Chapter 3).

Other techniques for eliciting reactions from participants, under set headings or totally free, are also described. In one such, participants are given three 5" x 3" 'post-it' slips towards the end of a teaching session or component with the instruction that one slip should be labelled STOP, the second START and the third CONTINUE. Participants note down some aspect of the course that they would like to see stopped, started or continued and stick their slips up on the board, wall or door in the area designated for STOP, START or CONTINUE, as indicated in Figure 2.4, below:

In this example, participants are then given a chance to look at what others have written and the slips are then taken down and processed by the tutor. This system is economical, attempts to elicit new ideas as well as reactions to current practice and seems likely to yield useful feedback on an individual level. One drawback of the technique is that the opinions expressed might relate to a very wide range of topics. Moreover, since there is scant opportunity for individuals to react to what others have written it is unclear how widespread opinions might be. Any solution to these problems would involve the specification of topics and more time, which have their own disadvantages.

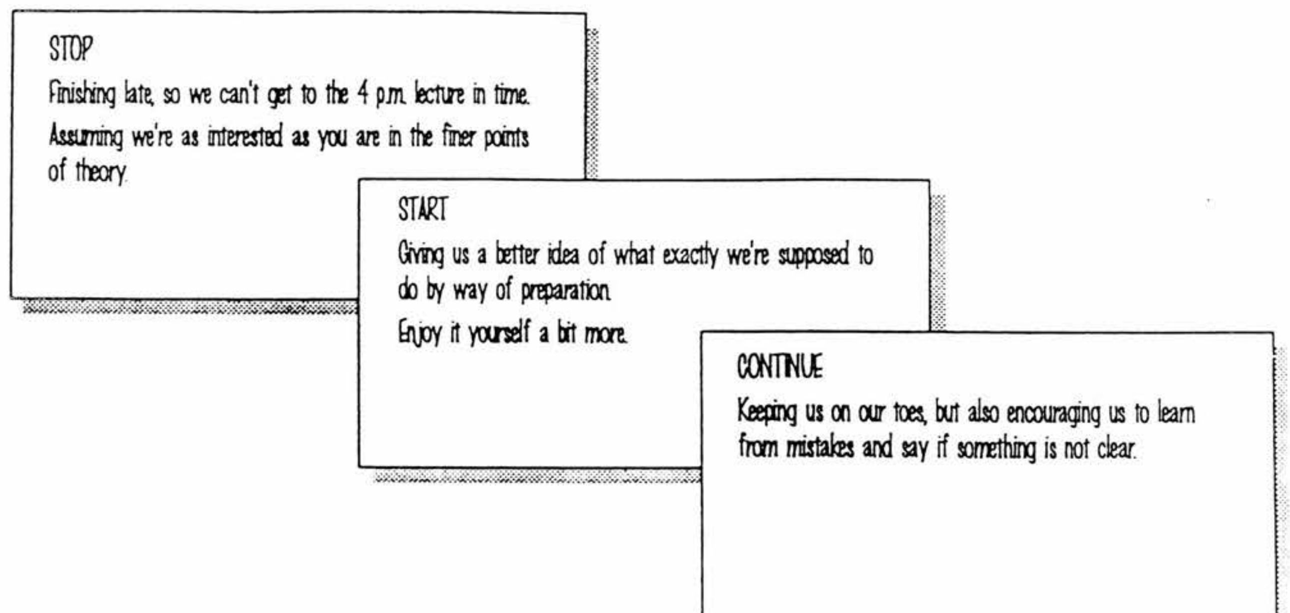


Figure 2.4: An example of 'post-it' feedback about tutorials
(TLA, Edinburgh University 1996)

Morrow and Schocker (1993) describe a more refined, but more time-consuming approach, the 'open poster forum' (1993: 50). Following an initial analysis by the evaluator of issues raised during individual interviews in relation to specific sessions, five posters were pinned up, four with headings summarising the feedback and one blank. Participants were asked to write comments on slips of paper and attach these to the relevant poster(s). They were then encouraged to read the comments on the various posters and discuss their reactions to these. The evaluator and course tutors absented themselves from this stage, but the written comments were subsequently collated by the evaluator and then presented to the course team. Plenary review sessions were also held on Friday afternoons at 4.30. Despite the timing of these sessions and the fact that they were optional, nearly all participants attended. By contrast, the institutional questionnaires that participants were asked to fill in on the final day of the course were completed cursorily and the course team 'had great difficulty in collecting them all in' (*op. cit.*: 54). One explanation would be that participants felt that by this point they had said everything they wished to say; Morrow and Schocker (*ibid.*) prefer another explanation:

> summative questionnaires ask the participant to judge the worth of the product which he or she has received: process evaluation as described in this paper invites the participant to share in the design of the product and to reflect on how it is made.

Feedback need not be confined to what participants say. It can also be inferred from what they do or know (performance on tasks). An interesting technique for examining the structure of teachers' knowledge and the effects of a course is 'concept mapping':

> Students are asked to brainstorm about a particular topic, such as planning or classroom management. Having produced a list of vocabulary or 'concepts' they are then asked to arrange them to demonstrate their relationships and interconnections and sometimes to label what these relationships are. The resulting concept map is taken to represent their understanding of the topic. ... It can reveal gaps, misconceptualisations, as well as the degree of sophistication of understanding as revealed in the connections that are made'
>
> (Calderhead 1990: 157)

As with the other forms of feedback referred to above, concept maps provide clear pointers to the necessity or otherwise of rethinking aspects of course provision.

Participants are not the only source of feedback, of course. Tutors' own views on aspects of course organisation, structure and content will be conveyed during formal or informal meetings. A Course Director or external evaluator may also wish to sit in on sessions with a view to forming an independent judgement of how these are conducted and participants' response to them.

2.6.4 End-of-course evaluation

It is common practice to elicit participant feedback at the end of a course. Although this cannot affect the course in question or its participants, it can have an influence on future courses. Whether this is actually the case depends on what is done with that feedback. Davis (1990) reports an incident during an end-of-course tutors' meeting when a novice tutor mentioned that one of the participants on the (CTEFLA) course had felt that more time should be spent on language analysis. The chairperson 'dismissed the contribution with the remark, "They always say that", and moved on, with nobody demurring, to the next item' (1990: 21). Davis says ironically, 'The observation was commonplace and so deserved no attention' (*ibid.*) and adds: 'Though this incident may not be typical, it does illustrate that ... results that conflict with the views prevailing within a training institute can be ignored, while suggestions which confirm particular prejudices can be adopted' (*ibid.*). Davis is arguing for more rigorous scrutiny of internal evaluation procedures by an examinations board, but his general point has wider relevance: that institutions need to ensure that mechanisms exist for the systematic collection and thoughtful assessment of reports and other potential input to programme evaluation. One control which might make a difference would be for participant feedback sessions to be conducted by a tutor who has not worked on the course under discussion (Davis, *op.cit.*), a point made in 2.5, above.

Most commentators are agreed that the results of end-of-course evaluation may be misleading (e.g. Davis 1990; Woodward 1991). Again, Davis's comments on short pre-service courses seem broadly applicable to INSET courses, especially where assessment is involved:

> Trainees are, in my experience, rarely in the best frame of mind to undertake a cool, level-headed appraisal of what for many has been a gruelling, if stimulating experience. Many are tired, some are exhilarated, some may be depressed or disappointed, a few will be preoccupied with their final teaching practice that afternoon, and most are at least concerned as to what grade they will receive on the course; one or two may have already written themselves off as having failed the course. Secondly, the feedback session takes place before any of the trainees have had an opportunity to experiment with the ideas and techniques presented to them on the course, except in the unrealistic context of teaching practice.
>
> (Davis 1990: 21)

The points made here by Davis, that affective and other factors render participants' feedback at this point unreliable, and that in the case of courses which have for a period of time taken teachers out of the classroom the practical value of the course cannot be assessed either, constitute a powerful argument for delayed feedback.

2.6.5 Post-course evaluation

The argument for post-course evaluation is that certain types of effect cannot be validly assessed until some time has passed. If participants have been removed from their teaching context during the course, they need a period during which they can come to terms with new ways of thinking and working, a period during which 'pressure and influence from those involved [in teaching them] have waned' (Woodward 1991: 221). They also need time for the 'more considered distillation of good and bad effects' (*ibid.*).

Without some form of post-course evaluation, course providers are simply trusting to hope. Breen, Candlin, Dam and Gabrielsen (1989) describe an INSET programme for English language teachers in Denmark which started as a one-off intensive short course concerned with materials for communicative language teaching but evolved over some years in response to feedback from participants and local tutors, who led periodic in-country workshops. The authors trace three phases in this evolution: in Phase 1, the focus was on training as transmission; in Phase 2, on training as problem solving; and in Phase 3, on training as classroom decision making and investigation. They make the point that if Breen and Candlin had not been invited back after the first course, they would not have been aware of the problems inherent in the approach adopted in the first phase.

A similar point is made by Alexander (1980) in an article on the evolution of advanced in-service courses for British teachers. Alexander also draws attention to 'the elusiveness of the goals which lie at the heart of advanced study' (*op.cit.*: 192). These goals are elusive because 'they are concerned with qualities of mind which we trust will open up and make as considered as possible the conceptual basis of a teacher's day-to-day practice' (*ibid.*). The conceptual and organisational difficulty involved in post-course evaluation may, as Alexander admits, act as a deterrent, but it is nevertheless important to obtain information on outcomes, even if some methodological compromises have to be made.

As indicated in Chapter 1, follow up is particularly difficult in the case of UK courses for individual overseas teachers or one-off courses conducted overseas by UK-based personnel. Attempts to carry out such evaluations are, therefore, both rare and valuable. Lamb (1995) conducted a two-week course in the teaching of reading skills for 16 Indonesian teachers in tertiary-level institutions. Returning a year later, he interviewed 12 of the teachers and observed four, finding that 'very few of the ideas presented on the course were taken up in a way anticipated by the tutors, mainly due to the mediating effects of the participants' own beliefs about teaching and learning' (1995: 72). His conclusion is similar to that reached by Breen *et al.* (1989): that participants on INSET courses should first be induced to articulate their own beliefs and practices, considering any potential contradictions between these, and then 'formulate their own agendas for change in the classroom' (Lamb, *op. cit.*: 79).

Davis (1990) writing of changes in the behaviour of pre-service trainees over the period of an intensive one-month course, comments that 'there is no guarantee that the changes in behaviour represent genuine modifications in attitude *or that such attitude changes as do take place will be permanent'* (*op. cit.*: 14, emphasis added), implying that longer-term investigation is needed. Henderson (1978) reports a number of studies of attitude-change in teachers following INSET courses. In several of these, positive effects noted during the course subsequently disappeared; in one study of teachers' dogmatism (Weis *et al* 1974) teachers actually became more dogmatic, when the opposite effect was anticipated. Henderson's general conclusion is that 'the attitude changes found in in-service courses are less wide-ranging than those often found during initial training courses ... and no more permanent' (*op.cit.*: 145). Citing McLeish (1969), he adds that one reason for the less

marked changes in teachers on INSET courses may be that the personalities, attitudes and educational values of the latter are more firmly established.

Alderson (1985a), having lightly caricatured some of the reasons for not conducting post-course evaluation (participants will not come back even if the course was a success; and if it was not a success, then there is something wrong with the participants), examines the underlying causes for failure to follow up. Where a course will not be repeated, there is understandably little practical incentive for an institution to invest time and money in investigating its effects. Even where a course will be repeated, the decision to make this investment requires forward planning - since it needs to be built into the course budget and affects the allocation of staff time after the course - and a commitment to make use of the results.

He goes on to describe the methods of follow up that have been used by the Institute for English Language Education at Lancaster University (implementation plans, questionnaire, face-to-face contact with former students and their colleagues - whose opinions were 'solicited informally on an opportunity basis' (140)) - and gestures towards other sources of information, such as sponsors, inspectors and those taught by former participants, none of whose views have been sought systematically.

The paper concludes with a consideration of some of the problems associated with follow-up evaluation, categorised by Alderson as logistical, methodological and interpretational. Problems of logistics are particularly associated with time and money, the time required to carry out the evaluation and the time needed by former participants to respond; investigations will also be more difficult if ex-participants are distributed over a large area, even if within the same country. Methodological problems are related to logistical problems. It may be that the only practical way of obtaining feedback from certain individuals is through questionnaire or letter, but even allowing for low response rates due to 'local busyness' (136) these have certain limitations. Moreover, there is the difficulty of knowing - in the absence of other data, and without the possibility of probing further - what value to attach to participant responses and, indeed, of deciding which responses to attend to and which to ignore. This issue of evaluator bias (the tendency to highlight those responses which confirm one's own views, or deliberately to suppress these, as over-compensation) applies equally, as Alderson notes, to ongoing course evaluation. A final

problem concerns implementation. It is sometimes difficult to establish whether the problems reported by ex-participants in the course of post-course evaluation are, as Alderson puts it, due to the course or despite it, and to know, therefore, what action to take. The ultimate answer may lie in preparing course participants to cope with their problems independently; however, one paradox of Alderson's paper is that the follow-up evaluations he describes revealed that the UK courses (or perhaps the follow up itself) did little or nothing to reduce dependency on outside 'experts'.

2.6.6 Informal evaluation

Formal evaluation entails the planned collection of information; informal evaluation (or what Rowntree (1985: 267) calls 'casual evaluation') is less systematic both in its intentions and the use made of the results. During a course, for instance, a tutor listening in to discussions between participants may gain some insight into what they already know, have gained from the course or still need to work on; if the course involves participants teaching, then the tutor can also see how ideas presented on the course are adapted and with what effect, and can therefore learn something about the trainees, the exploitability of the idea and his/her own presentation of the idea (Woodward 1991). This kind of listening-in can even form part of the course design. McGrath and Altay (1990) describe a simulation activity in which participants play out meetings between staff and 'problem students', and tutors move from group to group monitoring but not intervening in the discussion. On a formal level, the activity provides practice in dealing with 'typical' problem participants and participant problems and a basis for analysis of both problems and their solutions; however, it is deliberately placed at a point roughly half-way through the course, when genuine problems may well be simmering beneath the surface, and therefore offers an outlet for these. Other more familiar examples of informal in-course monitoring would include an alertness on the tutor's part to interpersonal relationships, which might affect the productiveness of pairwork and groupwork, and an awareness that apparently casual remarks made by participants during a coffee-break, about another tutor or the pressure of work or their difficulty in getting hold of books, may be less casual than they seem and necessitate further investigation. After a course, incidental contact with previous participants (in person, letters and cards) and comments from those who know them, may also produce evidence of lasting effects, affective or otherwise.

It is natural that more weight should be given to the information collected through formal evaluation, but incidental bits and pieces of information of the sort referred to above may, if laid out and grouped, also start to assume meaningful patterns. Chapter 3 puts the case for participant journals as a programme evaluation instrument; a case can also be made for a Course Director keeping a daily log (Rowntree 1985) and for tutors keeping journals (McDonough 1994). Analysed from time to time, any (and preferably all) of these may reveal connections and contradictions that would not have been noticed in the normal run of events and call for short-term action; at the very least, they would provide a sensitive record of the life of a course that might have a formative influence on future planning.

2.7 CHOOSING A METHOD

2.7.1 Introduction

One major limitation in the books on educational research and evaluation methods consulted for the present work - and these include a number of well-known 'basic' texts (Borg and Gall 1989; Bell 1993; Cohen and Manion 1994) is that they give little or no guidance on the choice of methods. To put it more precisely, they provide no general criteria by which methods can be compared and an appropriate selection made for a particular purpose. Bell (1993), for example, in *Doing Your Research Project*, a 'guide for first-time researchers' and an Open University set text, gives the commonplace advice that methods chosen should be those most likely to provide answers to the questions the researcher wishes to ask. This may be unobjectionable but it is hardly helpful, since any one of a number of methods (or combination of methods) might conceivably prove suitable in these terms. There are, it is true, occasional tables in such texts in which one method is compared with another, but the general approach is to treat each method separately and consider its *pros* and *cons* in what amounts to a methodological vacuum. This approach may be adequate for an experienced researcher-evaluator who is familiar with the range of options available; it is of less value to the inexperienced evaluator who wishes to be clear what the options are before making a choice.

Weir and Roberts (1994) offer a little more help in that they give an illustration of method-focus match. Table 2.1, below, is a version of that cited in Weir and Roberts (*op.cit.*: 135) and taken from Siedow, Memory and Bristow (1985: 138). The intention is to indicate

which methods can be used to evaluate each of the three 'objective types': teacher beliefs, teacher abilities and teacher practices. In itself, this is a laudable endeavour and one that ought to represent progress beyond that offered in the large mass of the books surveyed. Unfortunately, the resulting table offers little more than an indication that all of the four methods (questionnaire, interview, observation and document analysis) appear broadly speaking to be equally well suited to the evaluation of each of the objectives.

Table 2.1: Methods for evaluating various objective types (based on Siedow *et al.* 1985: 138

	QUESTIONNAIRE	INTERVIEW	OBSERVATION	DOCUMENT ANALYSIS
TEACHER BELIEFS	pre-/post	pre-/post	pre-/post	pre-/post lesson plans
TEACHER ABILITIES	pre-/post	pre-/post	pre-post (including videotapes)	pre-/post lesson plans; self-assessment quizzes
TEACHER PRACTICES	YES	YES	YES (incl. videotapes)	lesson plans; record of activities

Weir and Roberts (*op.cit.*) place the table (which in the original also includes 'student behaviors' and 'student learnings' and is in a different format) without comment alongside a commentary on their approach to method-focus match in their Nepal INSET study (1994). Unlike the table, the commentary helpfully illustrates some of the factors and choices involved:

> To measure student learning gains (the key summative indicator required by the funding agency) language test scores were obtained and compared. In order to get some direct evidence on the implementation of training in the classroom, observational data were needed. To complement these direct data, teachers' unstructured self-report accounts of their lessons were also needed because resources did not allow a large enough number of observations. Focused interviews were carried out too because biographical data about the teachers in the study were needed.
>
> (Weir and Roberts 1994: 134)

They also comment on the interaction between method-choice and other factors, such as real-world constraints:

> the choice of self-report method can be determined by a combination of resource constraints (sample size and access) and the nature of the information sought. For example, in a case where large samples are necessary, or personal access is a problem because of travel or time constraints, then postal questionnaires would be an appropriate means of data collection. However, postal questionnaires (completed by informants on their own) are most suited to low-inference, *factual* questions - for example, information on conditions of service, or target situation needs. On the other hand, high-inference information (such as teachers' levels of expectation, reactions to a new course book, or students' perceptions of what they have been taught) is more suitably gathered by interviewing because it affords greater scope for questions of complexity and depth. Interviews are particularly useful where explanations of behaviour or affective responses are needed and as a means to pursue pedagogic issues in depth.
>
> (*ibid.*, original emphasis)

The constraints, they explain, may include logistical factors (access to informants, the adequacy of communications, and the availability of resources); the characteristics of informants (this point is not glossed) will, they say, be a further consideration.

The next subsection extends the references to logistical factors in Weir and Roberts and incorporates these within a tentative framework for the comparison of methods. The framework is then applied in Chapter 3 to the description and analysis of specific methods.

2.7.2 A framework for the comparison of evaluation methods

Mention has been made above of one of three major considerations that are likely to influence the choice of method: logistical factors. The second major influence is what for want of a better term might be called 'methodological factors' and the third is utility.

2.7.2.1 Logistical factors

The most significant logistical factors are the following:

ñ economy;

ñ availability of essential resources;

ñ ease of administration (including access to informants and ease of communication).

Economy

The concept of economy covers both actual monetary expenditure and time. Expenditure might be entailed in respect of postage (for postal questionnaires) or telephone calls (for telephone interviews), travel (for observation), fees (for an external evaluator, for instance) or training time for personnel. There could also be costs in relation to equipment (the purchase of audiotapes and videotapes, and the copying and editing of these). Time is, however, the largest potential expense: time for the preparation of data-collection instruments (e.g. interview schedule, questionnaire, observation coding system) and for the piloting of instruments; time for data collection and processing (potentially a great deal of time if transcription is involved); time for data analysis; and time for report-writing. The time required of the respondent, although not a cost in the same sense, will also be a consideration, a relevant factor here being the extent to which the individual's goodwill is involved.

Availability of essential resources

Certain types of evaluation activity are equipment-dependent. An evaluator might wish to make use of one or more of the following: cassette-recorder, preferably portable and preferably battery-operated; microphone, possibly of a type and quality capable of being used for classroom recording; video camera; computer; and have access to reprographic facilities and copying/editing equipment. Skilled personnel might also be needed to operate such equipment or to collect and process data of particular types. If certain items of equipment or skilled personnel are not available, and there is no budget for the purchase of new equipment or the training of existing personnel, this will obviously constrain what is possible.

Ease of administration

The ease or otherwise of administration is in part a product of the scale of the evaluation. If this is conducted on-site and involves a relatively small number of subjects, the problems will be limited to finding a suitable time, a suitable room or rooms and the necessary personnel. For both on-site and off-site evaluation, however, equipment may be a concern. In many contexts it is no easy matter, for instance, to transport the equipment necessary for videorecording classes. Overseas, this can be a particular problem in rural areas -

even if there are suitable conditions for filming, access to informants will be problematic if the infrastructure (transport, postal service, telephone system) is unreliable.

2.7.2.2 Methodological factors

The most important methodological factors are ease of analysis, objectivity, validity, reliability, comparability and generalisability of results. The issue of propriety is dealt with separately in section 2.6.5, *Politics, ethics and communication*. For purposes of evaluation, these methodological factors can be considered under two main heads:

- reliability (together with other issues related to this) and validity
- comparability and generalisability.

Reliability and validity

In assessing reliability we are concerned with consistency and objectivity. We need to be as sure as we can be that the performance (e.g. interview response, lesson) if elicited on another occasion would be substantially the same; and we need to feel confident that the judgements made about that performance are replicable and as free from subjectivity as possible. The central question in relation to reliability, as the term suggests, is whether we can rely on the data and results.

Within INSET evaluation, reliability is affected by a number of factors, some of which will also be applicable to any kind of evaluation. These include the possibility of subjective judgements (e.g. where interview or observation data has to be classified according to high-inference categories) or where assumptions have to be made about the representative nature of the data for the person or persons generating that data. Since objective data places fewer demands on the evaluator, being easier to analyse and quantify, it is less subject to differences of interpretation and therefore inherently more reliable. One limitation of the kind of data which might be deemed objective, however, is that it may also be rather uninteresting in isolation. To know that Teacher X performs a particular classroom act *n* times and with a particular distribution over the space of 40 minutes may tell us something about Teacher X and may allow us to compare Teacher X with Teacher Y, but it does not get us very far if we wish to understand why the action was performed more frequently during one segment of the lesson than another or to know with what

degree of skill the action was performed. Important though it is, excessive insistence on reliability might be trivialising. On a more positive note, the concern for reliability causes evaluators to take precautions against predictable bias, in such forms as interviewer or observer effect and 'observer drift' (the tendency for coders to deviate over time from an agreed standard), one consequence being that results are more likely to be taken seriously within academic circles.

As Bell (1993: 65) notes, validity 'is an altogether more complex concept' [than reliability]. Essentially, it is concerned with the representation of reality, as reflected for example in the extent to which a sample of informants (e.g. those selected for interview) represents the whole population, and the relevance, completeness and accuracy of the data (i.e. the extent to which the data is uncontaminated by collection procedures or informant misrepresentation, and reliable). As the previous sentence indicates, data must be reliable if it is to be valid; but reliability in itself does not assure validity. (For instance, informants may give the same responses or produce the same behaviours on a number of occasions and for different investigators, but what they say or do is not necessarily a true indication of what they really think or would do under normal circumstances.) Technically speaking, data may have internal validity if it satisfies certain criteria relating to the way in which it is collected, but we are unlikely to set much store by it if it does not also match our expectations and intuitions or is not supported by other evidence.

One way of representing the relationship between validity and reliability is shown in Figure 2.5, below.

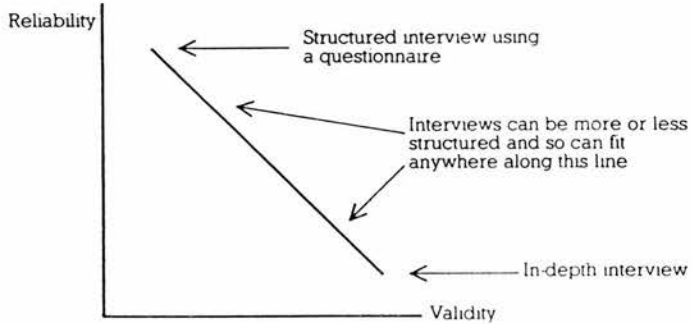


Figure 2.5: The reliability and validity of interview data (Langley 1987)

Either extreme involves sacrifice. To opt for maximum reliability involves loss of validity, and *vice-versa*. At the heart of the tension is data. The most reliable data, in terms of judge reliability, is that which is easy to analyse and compute. The most valid data, valid in that it takes the form of relatively free expression, is by contrast more difficult to manage. It can be managed, however, as Walker (1985: 185) testifies: 'analysis involves "fracturing" data into "lumps of meaning" (e.g. events, actions, acts, statements, concepts) and a subsequent restructuring, first by categorisation and then by developing relationships between categories'. It may even be possible subsequently to convert the categorised data into quantified results. Such processing inevitably involves a series of more or less subjective decisions, rendering the results at each stage increasingly less reliable.

In the testing literature, five types of validity are normally distinguished (face validity, content validity, construct validity, concurrent validity, predictive validity), each of which has potential relevance to programme evaluation. *Face validity* is the extent to which an instrument (e.g. questionnaire, observation checklist) appears to measure what it is supposed to. This is important for what might be thought of as 'public relations' reasons: an instrument which lacks face validity may negatively influence the attitudes of those on whom it is to be used or those asked to administer it or sanction its use. These attitudes may extend beyond a reluctance to cooperate to more actively manifested distrust of the whole endeavour and of the evaluator. In this sense, the latter may be judged by the quality of the instruments he or she proposes to use. The assessment of face validity, which is concerned with the more obvious, superficial aspects of an instrument, is subjective. *Content validity*, on the other hand, is assessed more objectively, by means of the careful study of, for example, the relation between a teaching programme and a test based on that programme. A test which has content validity will deal with a representative sample of the programme and do so in a way which appropriately reflects the balance and types of activities involved. Similar strictures apply to the questions used in questionnaires and interviews and to the way in which survey samples are constituted. On award bearing programmes, it may be appropriate to make judgements of the *predictive* or *concurrent validity* of an assessment procedure. These rely on comparisons of at least two sets of results: an instrument is said to have predictive validity if it is subsequently shown (by other measures) to be an accurate predictor of what is taken to be the reality; an instrument has concurrent validity if other measures used at the same time produce similar results. Finally, c*onstruct validity* is concerned with the adequacy with which theoretical constructs

are conceptualised in the construction of an instrument. It thus bears a resemblance to face validity, but whereas construct validity rests on a foundation of analysis and logical development, face validity may prove to be no more than skin-deep.

A further type of validity, *descriptive validity*, is suggested by Runciman (1983), cited in Walker (1985: 190). This is best illustrated by reference to its converse, 'misdescription', which takes the form of 'incompleteness, oversimplification, suppression, exaggeration and ethnocentricity' (*ibid.*). Incompleteness, explains Runciman, is neglect of issues which may be of significance to the respondent but are peripheral to the researcher; oversimplification results from the failure to ask for description in the respondent's own terms; ethnocentricity is reflected in the inappropriate mapping of the researcher's assumptions on those of the respondent - particularly important, perhaps, in the case of cross-cultural interaction; in contrast to these, suppression and exaggeration may be a deliberate tactic on the part of the researcher, since they suit his/her own purposes.

Comparability and generalisability

One of the expected outcomes of scientific enquiry is that the results obtained from one subject (or study) will be capable of comparison with those obtained from another subject (or study) and that statistically significant tendencies will yield generalisations. This may be possible in circumstances where all variables can be controlled; but is much more difficult in educational settings where different groups of participants are involved, and where ethical and financial considerations might constrain experimentation. The problem of comparability is, of course, particularly acute when the data generated by an evaluation study is qualitative. Although it is possible, as noted above, to scan such data for patterns, categorise it, and thereafter quantify it, this is a recursive, lengthy process and one in which there is only too evidently subjectivity and scope for error, with the consequence that the results, in scientific terms, have suspect reliability. For this reason, it is desirable that evaluation studies provide corroborative evidence either through triangulation of method or triangulation of source.

2.7.2.3 Utility

If the academic community requires that evaluation findings be reliable and valid, stakeholders in addition expect the findings to be useful, to offer answers to their questions. Unlike many other forms of research (see the discussion in Chapter 1),

evaluation is practically-oriented, one of its objectives normally being to inform decision-making for developmental and/or accountability purposes. Indeed, one view holds that the value of evaluation lies in its contribution to policy-making. From this perspective, evaluators put their technical expertise at the service of stakeholders, are bound by stakeholders' agendas of concerns and questions, and are judged by the relevance of the answers they provide. This in essence is what has come to be known as 'utility-focused evaluation' (Patton 1986, 1990).

A version of the same principle might be applied to the comparison of methods of enquiry. In selecting one method rather than another, evaluators will be influenced by the factors discussed above but they will also be interested in the particular contribution of a method to the questions they wish to answer. It is perhaps this understanding that lies behind the bland injunction to choose a method that seems likely to provide answers to the questions one has in mind. In Chapter 3 an attempt is made to identify the specific contribution of each of the methods surveyed.

2.7.2.4 Conflict between principles

In certain situations, the principle of utility, for instance, might appear to favour the choice of one method over another, but this choice might prove to be wholly inappropriate when the principle of economy is applied. The fact is that some degree of compromise between economy and other principles will often be necessary. When the compromise is on the side of reliability or validity, it is important that this is recognised and the results treated with particular caution.

A recent report from China (Ward, Barr, Chai, Hua, Kong, Lu 1995) of an evaluation of two INSET programmes in which a range of off-site measures were used shows the kinds of compromise that typically have to be made between validity and reliability on the one hand and economy on the other. The researchers were interested to establish the effects of the programmes on participants' normal teaching practices. In addition to participant-focussed lesson recordings and interviews, recordings were also made in relation to one of the programmes of the lessons of participants' colleagues; students were interviewed in an attempt to gain some insight into the situational constraints; and a questionnaire based on all the information gathered was subsequently sent to all former participants. However, as frequently happens, the economy principle also played a role, in this case affecting the

selection of candidates for face-to-face interview and recording and the number of lessons that could be observed.

The point being made here is not that the principle of economy should be subservient to that of validity, say, but rather that the two should be in proper balance. If the economy principle is allowed to determine the choice of method then the resulting data may be too limited or too unrepresentative to serve any useful purpose; if the validity principle is unrestrained, then evaluators may drown in data. As Elley (1989) puts it, there is no point in using a sledgehammer to crack a nut. This point applies with even more force to evaluators operating on a part-time basis (the normal situation within teaching institutions), who need methods that are 'realistic and economical' (Weir and Roberts 1994: 31). Compromise and adaptation may be necessary, Weir and Roberts concede, thinking presumably of sample size, but not at the expense of systematicity of approach nor the accuracy and quality of the information gathered. For many evaluators, in many institutions, these conditions may be very difficult to satisfy.

2.7.3 Constraints

It is difficult to imagine an evaluation without logistical or methodological constraints. For institutions, competing demands on staff time are likely to constrain data collection and processing both during and after the event, although a commitment on the part of management and the enthusiasm of individual members of staff could make a difference. Within smaller institutions limited expertise might be a further constraining factor. One consequence of such logistical constraints, particularly on short courses, is that compromises tend to be made at a methodological level, the most obvious of these being reliance on a single method of enquiry at one point in time (e.g. the end-of-course questionnaire). Even where institutions make a more determined attempt to obtain post-course data, ease of access to participants and low return rates for questionnaires (with responses coming only from the more positive, perhaps) militate against strong conclusions.

2.7.4 Politics, ethics and communication

2.7.4.1 Introduction

Evaluation of courses involves, directly or indirectly, evaluation of people. This can be a particular problem in external, accountability-driven evaluation and when, as in projects, perceptions of evaluation may differ sharply between expatriate and local personnel.

Murkowska, for instance, writes tellingly of the background to attitudes to evaluation in Poland, where it was 'a threatening means of control' (Kiely and Murkowska 1994: 67) and might lead to the dismissal of staff and the closure of programmes and 'ideologically hostile' departments (*ibid.*). Anxiety on the part of programme personnel was accompanied by cynicism:

> even when evaluation was carried out, its purpose was either to provide the authorities with the data supporting desired outcomes, or if it happened to be carried out more objectively, it was a type of summative evaluation, focusing on the overall outcomes used for the statistics. Never ever has it had a developmental function, which could serve teacher self-development and curriculum development.
>
> (*ibid.*)

It will be clear from the above quotation that negative experiences deriving from one kind of evaluation may colour attitudes towards evaluation in general unless they are recognised and dealt with. This has implications not just for expatriate project staff but also for UK institutions receiving overseas students. Suspicion and cynicism hardly form an appropriate basis for positive cooperation or openness. The problem is not simply one of cultural difference, of course. Alexander, referring to INSET evaluation in the UK, also warns of resistance to judgements by others. 'Evaluation "from above" can be used to inform decisions about teachers as well as courses; peer evaluation can engender a profound insecurity; and, for many, evaluation "from below" may seem impertinent' (1980: 184).

In Alexander's hierarchy, those "below" are pupils; viewed from the position of a teacher educator they might equally be the teacher-participants on an INSET course. Both of these learner groups may feel equally uneasy when asked for their comments on a course. Overseas teachers especially may be unwilling to voice direct criticism because this reflects

on those in positions of authority - a standpoint motivated by personal sensitivity or cultural conditioning. Sponsored students or those following award bearing courses may also be concerned that they are themselves being evaluated in some devious way and that critical comments will be used against them (Sharp 1990; Davis 1990).

At the root of the problem are three interrelated issues: politics (the locus of control and how power is used), ethics and communication.

2.7.4.2 Politics

It might be assumed that the political dimension of evaluation is most in evidence at the level of externally motivated evaluation where the results are used to promote or consolidate particular positions. It will also, of course, manifest itself in the nature of the relationship between the professional evaluator and the contractor (Macdonald 1976; Stake 1976). Interestingly, the latter kind of relationship is not always one-sided. Elliott (1980), for instance, writes of his own tendency to criticise the higher orders of the particular slice of bureaucracy with which he happens to be dealing and to explain and act as advocate for the underdog.

To equate political action only with this level of evaluation would be mistaken. To quite a large extent, politics is about participation (see 2.4.1). The way in which decisions are taken at the planning or implementation stage of any evaluation can be construed in a political light (are all groups consulted ? are all party to decisions ?) and which decisions are taken (e.g. when the findings are known, what action is taken, what information is released/suppressed) (Weir and Roberts 1994). At the individual level, there may be effects in terms of self-esteem deriving from whether or not the individual is consulted and whether his/her practices and views are identifiable and the subject of confirmation/ disapproval (Weir and Roberts, *op.cit.*). As Alexander (1980: 83) puts it: 'Public, formal evaluation is about relationships and power distribution between individual groups and levels'. Where groups or individuals, including evaluators, feel that they have insufficient control over the decisions that affect them, the conduct of the evaluation and/or its results are likely to be affected.

2.7.4.3 Ethics

The question of ethics in ELT research has recently been raised by Dufon (1993), arguing for the revision of the TESOL Guidelines. TESOL, the American association of teachers of English to speakers of other languages, had in 1983 published for the first time 'Guidelines for Ethical Research'. The introduction to the guidelines makes the point that although local guidelines may exist, prepared by universities, school boards or funding agencies, for instance, these need to be supplemented because ESL learners represent a particularly disadvantaged group: not only are they on the wrong end of a power differential *vis-à-vis* the researcher, they may also be linguistically disadvantaged. In general, the guidelines were intended to safeguard the rights of learners in research involving human subjects: five of the six principles related to informed consent, deception, consequences in terms of risks to learners, privacy and confidentiality and anonymity. The sixth principle concerned the applications of research, where the anxiety was that a clear distinction be made between well-grounded findings and speculation, lest learners suffer from overenthusiastic consumers of research reports. Dufon suggests that changed circumstances (e.g. technological advances and data-sharing) justify a reconsideration of the guidelines. She also proposes that the concentration on learners could be widened to include teachers and programme administrators, who may also be a focus for research - a suggestion of particular relevance for evaluation.

In the context of programme evaluation, ethics is a major concern, and one that is widely recognised (see also the Guidelines of the British Association for Applied Linguistics (1995) and those of the Joint Committee on Standards for Educational Evaluation (reproduced in Stufflebeam 1990 and here in Appendix 2.3)). What is at issue is the protection of the rights of individuals (programme participants, programme personnel, and others) and, for that matter, institutions, to be treated fairly and consulted before their views are divulged to a wider public. This is not simply a question of explicit naming or seeking permission before quoting; neither individuals nor institutions should be identifiable unless they consent to this willingly.
One difficulty with this principle is its one-sided protection of the individual. In a thoughtful discussion of ethics in accountability-oriented evaluation, Parsons (1990) observes that there may well be a tension between the right of the individual or institution to be unidentifiable and the right of the sponsor to know if things are going badly and

where any problems are located. In another helpful contribution, Simons (1979) argues that pseudonyms or 'role designations' should be used on the basis not that they offer anonymity (which they may not) but because they depersonalise critical issues. In similar vein, she suggests that the criteria for reporting be 'fairness, accuracy and relevance, not ... whether the person looks favourable or unfavourable in the report' (1979: 130).

2.7.4.4 Communication

In the sphere of evaluation, communication serves a number of very specific functions. Some may be more relevant to externally-driven evaluation, but all also merit consideration by internal evaluators. The functions of communication include:

ñ negotiation between evaluators and stakeholders to establish terms of reference
ñ eliciting information necessary for the planning of both a communication strategy *vis-à-vis* stakeholders and an evaluation strategy
ñ eliciting data
ñ informing stakeholders of plans, progress, etc
ñ reporting findings.

By way of illustration, let us briefly consider just two of these functions: *negotiation* and *eliciting information as input to a communication strategy*. Issues relating to *reporting* are discussed under 2.8.

The importance of negotiation is stressed by Beretta (1990b), arguing for the independence and integrity of the evaluator; similarly, for Weir and Roberts (1994: 211) 'the defence of independence begins with the terms of reference negotiated with the sponsor'. Not all commentators would agree (see 2.5 on the role of the evaluator). The reality may be that evaluators whose livelihoods depend on client satisfaction are unwilling to jeopardise this by insisting on points of principle. Moreover, as Parsons (1990) recognises, the evaluator's freedom to negotiate is dependent on the contractor's willingness to negotiate; contractors may be decidedly reluctant if they can see the interests of other parties than themselves being served by an evaluator's proposals. None of this affects the fact that negotiation will normally be desirable, not least because the evaluator may be the person best placed to judge the feasibility of the terms of reference. The most convincing arguments may be found in examples of good and bad practice. Beretta (1990b), having referred disarmingly to his own failure to negotiate the research questions

underlying the evaluation of the Bangalore project (Beretta and Davies 1985), cites Rockwell's (1982) approach as exemplary. Rockwell, evaluating a programme designed to prevent alcohol abuse in young people, started with three questions of her own (Who is interested in the evaluation and why ? What decisions are to be made as a result of the evaluation ? What do you want to know about the project ?). These enabled her to identify the policy-shaping community and determine the research questions and an appropriate evaluation model. Within an institution, negotiation may also be necessary between the individual(s) carrying out the evaluation and those likely to be most affected, since the cooperation of the latter will be necessary.

Although negotiation may be the first stage in an extrinsically-driven evaluation, sponsors are clearly not the only stakeholders. Participants on a short INSET course which the course provider wishes to evaluate primarily for developmental purposes may be unwilling to give up much learning time for evaluation unless they are persuaded of the benefits, especially if they are paying for the course themselves. And if evaluation is imposed upon them willy-nilly, it may be resented. Some insight into their likely attitudes will therefore be important in determining a communication strategy. If evaluation of the courses they themselves teach seems a proper subject for discussion on the course they are following, this may be a convenient springboard; if not, some attempt can be made to elicit attitudes during informal conversation.

At the back of any evaluation there is an interesting issue of control. Who will decide, finally, whether the course will be evaluated and by what means (a question touched on in 2.5) ? In some contexts, the question of control amounts to one of ownership: if INSET is in some way concerned with self-directed professional development - a view that has been expressed in British educational circles - it could be argued that such decisions logically fall to participants themselves.

In the context of projects, the preliminary steps leading to the formulation of a communication strategy will necessitate consideration of such questions as whom to contact, through what medium, what should be said, as well as such interpersonal factors as form of address and tone. The importance of a well-judged communication strategy cannot be overestimated: communication keeps people involved (Rea-Dickins and Germaine 1992) and evaluators cannot operate effectively without full cooperation. Yet as

Holliday's (1992) discussion of formal and informal orders makes clear, what one is told as an outsider may for various reasons prove to be inaccurate, and it may be extremely difficult to collect accurate information. Institutions are microcosms, of course, and much of what constitutes good practice in communication within projects also holds good for communication within institutions.

2.8 REPORTING

The three issues of politics, ethics and communication come together in relation to reporting on evaluations. Parsons (1990) points out the obvious consequences of a report which fails to satisfy political requirements: 'Cause discomfort [to a sponsor] or give ammunition to others and your future evaluation services are not required' (148). However, it is the relationship between an evaluator and a sponsor of developmental evaluation that gives rise to particular tensions and is of particular relevance to institution-based evaluation by insiders. As Parsons asks wryly: 'Do you produce an evaluation report that damages your friends?' (*op.cit.*: 149).

The nature of the audience (specialist, non-specialist or a mixture of the two) should also affect decisions concerning the report (Hargreaves 1989). These decisions would relate to, for instance, the medium, the style adopted, the content (e.g. level of detail), and the discussion and presentation of results.

2.8.1 Style and content

It is generally agreed that a report should be clear and accessible to the audience for which it is intended. Beretta presents the pragmatic view: 'It would be unprofitable to adopt a discourse style that the audience will not or cannot listen to. After all, you can knock forever on a deaf man's door' (1990b: 10). Pragmatism also dictates judgement concerning brevity. Morris and Fitz-Gibbon (1978) recall the comment of a legislative assistant in America who was asked how evaluation could be rendered more useful to State Assembly members. The response was: "Just write down the conclusions of your report in one sentence, in large type, in the middle of a sheet of paper". They add: 'Influential reports are short and to the point' (1978: 10).

In broad terms, the content of a report will be affected by the purpose of an evaluation (Simons 1979) as well as its audience (Morris and Fitz-Gibbon 1978; McCabe 1980a; Ashworth 1985; Hargreaves 1989). Where there is more than one distinct audience or set of stakeholders - for example, the bureaucracy within an institution and programme personnel - it may be preferable to produce more than one version of a report, reflecting the information different groups of readers might want or need (Alderson 1992). If stakeholders have the kind of pragmatic interest ironically captured in the legislative assistant's comment, a report to them might be relatively brief and state its conclusions straightforwardly. Indeed, an 'executive summary' now forms an introductory section in many formal reports. Supporting this idea, Cooper offers his own law of evaluation: 'Faculty use of evaluation data is inversely proportional to the sophistication of the statistical analysis' (1983: 128). By contrast, a report written with one eye on the academic community - including other evaluators - would be likely to contain a rationale for the evaluation design, detailed information on procedures, and express conclusions in a tentative manner; statistical analyses of results, if these were appropriate, would be included in appendices. Alderson (1992) suggests that dates for action might also be incorporated, with the expectation that in due course a further report would be produced on the actions subsequently taken.

McCabe (1980a) lists the minimum contents of a report:

1. description of aims, physical circumstances and course programme; grouping, methods, outcomes
2. notes made/opinions expressed by course leader/tutors
3. opinions of course members
4. introduction, conclusion and structure provided by evaluator.

The emphasis in items 1-3 on description reflects McCabe's view that one of the roles of the evaluator is to register what happens. He recognises, however, that item 3 is the most difficult to deal with fairly (and interestingly and meaningfully), and that there is scope for bias in item 4. In its scope and reflectivity, this might nevertheless be an appropriate basis for reports primarily oriented towards programme development.

In contrast, Morris and Fitz-Gibbon's (1978) book 'How to Present an Evaluation Report', indicates the parameters that might be more appropriate for an accountability-oriented report. Chapter 2 of the book offers a 'standard outline' for a report that is 'intended to be exhaustive of the type of information than can be conveyed' (*op.cit.*: 7). Major sections are as follows:

1. background information on the programme: origin; goals; participants; materials; activities; administrative arrangements; staff
2. description of the evaluation study: purposes; evaluation design; outcome measures (instruments, data collection procedures); implementation measures (instruments, data collection procedures)
3. presentation of results: of outcome measurements; of implementation measurements; informal results
4. discussion of results: cause and effect relationships; programme effectiveness
5. costs and benefits (optional): method for calculating costs and benefits; costs of programme (financial; other); benefits of programme (financial; other)
6. conclusions and recommendations.

2.8.2 The role of the evaluator-reporter

Expectations of the evaluator as report-writer vary. Simons (1979), writing specifically of school-based self-evaluation, argues that the role of evaluators is:

> to describe what happens ... not to recommend what should happen, i.e. they should inform decisions without prejudging them; they should present options without prescription; they should come to no final judgement
>
> (Simons 1979: 129)

For Patton, who is intent on providing a vicarious experience for the reader, the report will contain description and quotation, these being 'the essential ingredients of qualitative inquiry' (1990: 429-30). Macdonald (1976, cited in Bolam 1980) similarly sees evaluation as a 'portrayal', while Beretta (1990b) characterises the evaluator as a storyteller. There is obviously a danger that storytellers become so engrossed in their stories that they forget their readers, and this will be a particular problem in case studies, where the writer is attempting to recreate the multidimensional reality of a specific context. Davies (1992), in a review of Alderson and Beretta's (1992) collection of case studies, is obviously a little

impatient with an approach which in the name of realism refuses to tidy up the messiness of research: 'An evaluation is not a history but an abstraction .... An evaluation must be an interpretation' (Davies 1992: 208). Judgement is necessary in the writing of any report, in particular judgement concerning what to omit. Patton (1990) puts this memorably: 'The agony of omitting on the part of an evaluator is matched only by the reader's agony in having to read those things which were not omitted - but should have been' (*op.cit.*: 429). Weir and Roberts (1994), in more po-faced fashion, note the importance of balance, the need for 'sufficient information to judge the reliability and validity of the procedures followed but ... a surfeit of information should be avoided' (1994: 129).

The real issue is perhaps the extent to which evaluators express their own interpretations, conclusions or recommendations. Should evaluators simply present a variety of interpretations for comment or judgement by others or should they present *their own* interpretations for judgement ? Saville and Andrews (in Rudduck, J. 1981: 70) make it clear that they prefer straight talking: 'we do not want to spend time grappling with a report in which the evaluator sits on the fence, and invites us to draw our own conclusions from the data and interpret the embedded meanings'. Descriptive detail may aid extrapolation, as Beretta (1990b) has observed, but those responsible for taking decisions based on an INSET evaluation will want something more explicit, such as the 'clearly drawn lessons' referred to by Weir and Roberts (1994: 139).

2.8.3 Circulation and participation

Different views are also expressed on the circulation of reports. While some commentators (e.g. Cumming 1988) see reports as potentially contributing to the field, others see them as confidential communications. Saville and Andrews, for example, in Rudduck, J. (1981) concede that although it may be appropriate to let participants see a report if they have contributed to it, they do not consider this - or wider readership - as particularly desirable. The principal reason, which they develop at some length, is that undue importance may be attached by the reader to the criticism which will inevitably be made in a report. Patton (1990), on the other hand, sees it as a point of principle to let interviewees see the draft report on their interviews and rewrites this in response to their comments. Alderson (1992), who shares Patton's view, points out that a report written from a single point of view, that of the evaluator, can subsequently be interpreted by each group of readers in

ways 'that suit their own particular interests or prejudices' (295). He therefore proposes that those concerned are shown a draft copy of the report and an attempt made to reconcile any differences (a form of triangulation at report-writing stage as well as at data-gathering stage). This should ensure that all perspectives and interpretations are reflected in the final report, thereby lending it greater validity. The approach espoused by Alderson, which also resembles Elliott's (1980) technique of trying out his hypotheses concerning the realities of a situation on those involved, can be justified on methodological, ethical and political grounds, but it is also eminently practical. As Alderson puts it: 'the reason many evaluation studies are not ... utilised is because one or more parties ... do not agree with the interpretation of events and results that have been presented publicly' (1992: 295-6). A potential drawback of the participative process, as Alderson (*op.cit.*) notes, is that some stakeholders may be diffident about expressing their views; other drawbacks relate to time - the time required of stakeholders and the timescale for the reporting phase. There may be little that one can do about the first two problems, but if realistic planning can make the difference between a report being utilised or not (for Patton 1990 and Alderson, *op.cit.*, the real test of its value) then it is clearly important for adequate time to be allowed.

For the evaluator, the report is the final stage in a process, yet it can also herald the start of a new cycle of activity in which the same set of interlocking decisions are involved, each necessitating consideration of a range of factors. The most complex among these decisions, as we have seen in 2.7, is that of method. This is the focus of the next chapter.

CHAPTER 3

DATA COLLECTION METHODS AND EVALUATION

3.1 INTRODUCTION

This chapter offers an overview of the methods typically used in educational and evaluation research and draws attention to other data-collection techniques that can also be used for this purpose. The intention is not to provide a comprehensive review of the literature but to describe the characteristic features of each method and, drawing on the criteria proposed in section 2.6.2 of the previous chapter, assess its potential relevance to INSET programme evaluation. The chapter therefore serves as a preview and critical background to Chapter 4, in which the procedures currently used in UK institutions to evaluate INSET programmes are surveyed.

3.1.1 Categorising data collection instruments

Festinger and Katz (1954, cited in Henderson 1978: 73) have noted that the social scientist has basically three ways of collecting data: by asking questions, by observing behaviour and by studying existing documents; Weir and Roberts (1994) similarly refer to asking, watching and reading. Henderson (*loc. cit.*) points out the variety possible within these broad categories (e.g. questions can be asked through questionnaires or interviews, and the latter may vary in their degree of structure) and the fact that in practice they may not be distinct (e.g. observation may prompt questioning and documents such as participant assignments may be a way of answering evaluation questions).

Within this tripartite categorisation subgrouping is also possible. Henderson (*op. cit.*), for instance, groups together interview and observation, both of which involve social interaction between the evaluator and participant. These, he holds, pose problems of skill and interpretation different from those required in the use of questionnaires. Since

documentary data is seldom used in isolation, he regards this simply as a supplement to other forms of evidence. Weir and Roberts (*op. cit.*) opt for another form of subgrouping, between self-report procedures (interview and questionnaire) on the one hand and observation on the other, basing this distinction on a difference in evaluation purpose (simply put, to establish what informants think, know, believe on the one hand and what they do on the other). In the interests of extending the range of possible method options in addition to the three basic methods of collecting data (questionnaire, interview, observation) and document analysis the chapter deals with what will be referred to as participant journals and participant plans, neither of which has received much attention in the L2 teacher education literature as a source of evaluation data. Since in the context of the present work there is no particular reason to adopt a more refined framework than that suggested by Festinger and Katz and Weir and Roberts, consideration of these six methods is sequenced in subsequent sections as indicated below, with cross reference as seems appropriate. Participant journals and participant plans, which as self-report instruments have more in common with questionnaires and interviews than document analysis, are included under 'Asking'. Section numbers are shown in brackets.

ASKING

Questionnaires (3.2)
Interviews (3.3)
Participant journals (3.4)
Participant plans (3.5)

WATCHING

Observation (3.6)

READING

Document analysis (3.7)

Although document analysis is not strictly speaking a data-collection instrument in the sense that data is specially elicited for the purpose of evaluation its inclusion is intended to highlight the relevance to programme evaluation of the documents that will normally be available. On award bearing programmes, these will include participant work and grades. The contribution of External Examiners to the evaluation of award bearing programmes is

referred to, but not explored in any depth since their involvement is a compulsory element and we are here concerned primarily with issues of choice.

3.1.2 The best method

Commentators are agreed that no one data collection procedure is inherently superior. What counts is fitness for purpose, i.e. what information is required, and - we might add - within a general framework of disciplined enquiry, what is the most economical way of collecting that information (Elley 1979).

In accountability-focused evaluation, that purpose will have been defined, with or without input from the evaluator, by the responsible person within the specific institution. Such evaluation is highly likely to include quantitative data, since this is what most managers expect, can grasp quickly (always assuming that it is presented appropriately) and can act on. A more developmentally-focused evaluation, one that is more concerned with processes than products, on the other hand, might contain relatively little quantitative data other than, say, tallies of responses on participant feedback sheets.

The general principle that the evaluation questions should determine the data collection instrument(s) employed is unexceptionable; however, at the level of implementation such logistical factors as access to informants and the resources available, as well as the characteristics of the informants themselves, may necessitate a degree of compromise, as Weir and Roberts are at pains to point out (*op. cit*: 132 and *passim*).

3.1.3 Combining methods

In many situations, the ideal solution will be to combine data-collection merthods in order to exploit their relative strengths. For example, Powney and Watts (1987) refer to two uses of interviews. In one, the interview is the first stage in two-phase research, serving as a qualitative pilot for the quantitative questionnaire stage. In the other, the interview is used after observation to clarify issues and opinions.

Henderson (1978), whose focus is the questionnaire, says that questionnaires may be at their most effective when they are combined with interviews. He suggests three possibilities:

1. An interview is used to discuss issues arising from questionnaire responses.
2. A questionnaire is used to ask relatively straightforward questions; an interview is then used for questions that are more difficult to frame or that may require a degree of probing.
3. The representativeness of responses given during selective interviews is checked by means of a questionnaire.

Weir and Roberts (*op.cit.*: 158) set out what they call a 'classic' sequence of stages in data collection which employs four distinct approaches to data-collection (document analysis, observation, interview, questionnaire) and a number of variations within the last two categories:

1. Examine all existing documentation (reports, records of attendance, teaching records, etc).
2. Do exploratory observation.
3. Do exploratory, unstructured or semi-structured interviews or group discussions.
4. Identify themes, issues, and topics emerging from stages 1-3 and use them as headings for the preliminary organisation of the data you have.
5. Write draft question items for structured interviews or questionnaires.
6. Pilot and trial a questionnaire.
7. Consider interviews as a validity check.
8. Administer the final questionnaire.
9. Analyse data.
10. Conduct follow-up interviews to illuminate or explain questionnaire response patterns.
11. Produce report.

As an example of a systematic approach to evaluation which draws on a range of data types and allows questions (or 'themes, issues, and topics') to emerge gradually, this is useful. It may be misleading, however, if it is seen as an obligatory series of steps or if the order of steps is understood to be fixed. In many situations, an evaluator will know which questions they wish to ask of the data and will choose that instrument or combination of instruments most likely to produce the required information economically.

3.1.4 Triangulation

Data collected by means of a single instrument may result in findings which are invalid because they are a product of the method, i.e. result in biased or distorted findings (Cohen and Manion 1994). One way of guarding against this possibility is to make use of what Cohen and Manion, following Denzin (1970), refer to as 'methodological triangulation' (*op.cit.*: 235) - that is, to base findings on data collected by two or more instruments. Methodological triangulation, which is the most common form of triangulation referred to in the literature consulted[11], has the further advantage over a single data-gathering procedure that it can be designed to produce both quantitative and qualitative data and thereby satisfy a wider number of potential interested parties.The less happily termed 'investigator triangulation' (Cohen and Manion, *ibid*), or triangulation of source, on the other hand, is an attempt to describe a phenomenon from two or more perspectives. In the context of INSET, these perspectives will be those of participants, tutors, and anyone else directly involved.

Triangulation is especially useful, according to Cohen and Manion, in the investigation of complex phenomena, where a holistic impression is required, and therefore lends itself well to case studies. It may also be useful, as a recent paper by Richards (1995) testifies, as a means of checking the subjective and possibly emotive reactions of course participants.

Richards had received adverse comments from participants on one component in a distance learning master's programme. He describes a series of measures to establish the nature and source of the problem: the standard written component evaluation forms were checked; interviews were arranged with seven individual participants; and twenty participants were asked to make detailed comments on two extracts (totalling seven pages) from the first unit of the component. The last of these measures was necessary, Richards notes, because 'interviews did not provide the best means of eliciting specific comments on the text itself' (*op. cit.*: 146). The assembled evidence indicated that further investigation was required, and he then set about close linguistic analysis of the text in question and a comparable text from another component which had not attracted any negative feedback. Interestingly, although his analysis eventually identified specific difficulties at the discoursal level, these did not correspond to the problems mentioned by participants - a reminder, if one is

needed, that in the search for objective reality we need as many reference points as possible.

3.1.5 Evaluating programme evaluation methods

In Chapter 2, the suggestion was made that criteria for the comparison of evaluation methods can be grouped under three heads: logistical factors (economy, availability of essential resources, and ease of administration); methodological factors (principally reliability, validity, comparability and generalisability of results); and utility.

These criteria are used to assess the instruments discussed in subsequent sections. The general aim of these sections is to indicate the options available to the evaluator and the factors that might influence decisions concerning evaluation method.

3.2 QUESTIONNAIRES

'responses can be misleading, unreliable, inaccurate and irritating'

(Rudduck 1981: 67)

3.2.1 Introduction

Questionnaires have been described as probably the most common way of collecting information (Youngman 1984) and, in the form of postal questionnaires, as frequently the best survey method for the purposes of educational investigations (Cohen and Manion 1994). Their outstanding characteristic is their convenience. As indicated in Chapter 2, they can be used at any stage in programme evaluation: as a way of gathering baseline data, typically prior to the commencement of a course; as a way of eliciting participant feedback during and at the end of a course; and some time after a course, as one source of information on effects.

This section outlines some of the choices open to the evaluator considering using a questionnaire and considers the strengths and weaknesses of questionnaires relative to other data collection instruments.

3.2.2 Questionnaire administration

Questionnaires may be:

1. read to the respondent: e.g. as a part of a structured interview (Peil 1982; Langley 1987);
2. given to the respondent for completion (variables being the degree of personal involvement by the evaluator, and where, when and under what conditions the questionnaire is completed);
3. sent to the respondent and returned by post (Langley 1987).

In the case of 2 and 3, anonymity will be an issue to be decided.

Decisions concerning which of these approaches to data collection is to be adopted will depend on such factors as the number, distribution and accessibility of respondents; the degree of formality of the study; and the relationship between evaluator and respondents.

3.2.3 Categorising questionnaires

Apart from the rather special use of a question schedule in structured interviews, questionnaires normally use written cues to elicit written responses. They vary in the type of cue and in the type of response the cue is intended to elicit.

Saville and Andrews (in Rudduck 1981) enumerate some of the techniques they have used to elicit participant feedback: asking respondents to underline statements that they agree with; answer specific questions; grade elements of the course; comment under specific headings; and, following an introductory briefing, write whatever they wish on a blank sheet. The last technique hardly qualifies as a questionnaire in the normal sense of the term, but what is interesting about the list is that its sequence reflects a gradual move away from structured elicitation in which both cue and response are highly specified. Saville and Andrews comment: 'we have learned ... that ... general cues can evoke rich data and allow course members to break away from the conditioning effect of the more traditional lists of questions' (*op.cit.*: 69). The disadvantage of 'rich data' from questionnaires or interviews is, of course, one of analysis.

Woodward (1991: 123-128) presents a variety of techniques for eliciting written student feedback, the majority of which are open-ended (see Appendix 3.1 for extracts). These include simple three-category open formats with the prompts *I liked, I didn't like* and *I suggest* (on which Vasconcelos 1994 - also included in Appendix 3.1 - offers variations), displays of participant comments (e.g. posters, a 'graffiti wall') and interactive techniques in which students react to tutors' questions and each other's responses to these. The Appendix also contains an example of a scale-based questionnaire.

3.2.4 Advantages and disadvantages of questionnaires

Table 3.1, below, presents the major advantages and disadvantages of questionnaires. As in subsequent tables, blanks indicate the absence of special advantages or disadvantages.

Table 3.1: Features of questionnaires

	ADVANTAGES	DISADVANTAGES
economy	large-scale data collection possible at low unit cost	time required initially for instrument development and piloting
resource requirements		design requires skill
ease of administration	administration requires no special skills; postal questionnaire convenient in countries with efficient postal service	
reliability	closed questions easy to analyse	
validity	internal cross-checks can attempt to establish consistency of responses; postal questionnaires allow time for consideration of response	return rate (and if large target population, sampling) crucial; responses may be incomplete, inaccurate (e.g. if dependent on memory) and untruthful, perhaps due to collusion or lack of anonymity; respondent may not understand concept or question; lack of flexibility (compared to interview)
comparability and generalisability	standardised, closed questions permit comparison of responses and, if return rate high enough (and sample representative), generalisation	generalisability dependent on return rate (and sampling)
utility	convenient means of collecting factual data which can be quantified and displayed by means of tables or graphs	limited value as a means of collecting information on attitudes

The table draws on Berdie and Anderson (1974), Henderson (1978), Youngman (1984), Langley (1987) and Weir and Roberts (1994).

As a data collection instrument, the main advantages of questionnaires over interviews or observation that are advanced in the literature are that they can be used to sample a large population relatively effortlessly and are convenient for statistical analysis. These may be factors if one is concerned to evaluate a large-scale INSET programme and if an evaluation is heavily weighted towards accountability, but they are not really a consideration for most institutionally-based programmes. Moreover, as Henderson (*op.cit.*) points out, interviews can also be used in a relatively economical and reliable way to sample smallish INSET populations. Where questionnaires do have an advantage over interviews for purposes of programme evaluation is that - in the shape of Rapid Assessment Forms (RAFs), for example - they can provide quick and easily quantifiable group-wide indications of levels of participant satisfaction with particular aspects of a course and therefore, in theory at least, permit modifications to be made from one day to the next. (Appendix 3.2 contains an example of a RAF from Rudduck 1981; for further discussion of RAFs, see Chapter 6.)

The major problems relating to questionnaires, and these are perhaps particularly relevant to questionnaires completed in the respondent's own time, concern validity. There may be a low return (which raises the issue of generalisability of findings) and the responses that are received cannot necessarily be taken at face value. This latter possibility can be attributed to one of at least two causes. The respondent may fail to understand a concept or question (Henderson (*op.cit.*) relates the personal anecdote of the teacher who, misunderstanding the instruction 'Ring your answers', tried to telephone him with the answers). Another possibility is that the response is incomplete or untruthful. Berdie and Anderson (*op.cit.*) note that one of the assumptions underlying the use of questionnaires is that respondents are willing and able to give truthful answers; this is not necessarily the case. As Peil says (and this is equally true of interviews):

> If asked something they do not know or have not thought about, most people will try to give some answer, just to be polite. It may be a guess, which they suppose to be true, or very wide of the truth to deliberately mislead the questioner, or it may be the first thing that comes to mind - to satisfy and get rid of the questioner.
>
> (Peil 1982: 100)

Similarly, questions which make demands on memory, knowledge or openness concerning sensitive issues may trigger a negative reaction. Slembrouck's (1987) analysis of

questionnaire responses highlights categories of response which are polite and face-saving on the one hand and aggressive and uncooperative on the other.

Slembrouck claims that 'the traditional questionnaire is noteworthy for its onesidedness of what is thought of as communicatively relevant' (*op. cit.*: 85) - in other words, researchers dictate what they wish to hear. He adds: 'Despite the anonymity contract, respondents are caught between "free expression" on the one hand and the demands of cooperative and socially acceptable and "appropriate" linguistic behaviour on the other' (*ibid.*). These comments raise three related issues, all of which affect validity. The first has to do with whether the questionnaire allows scope, within the framework laid down by the designer, for respondents to amplify their answers. Most questionnaire-designers are only too well aware of the importance of keeping a balance between informativity and the demands made on the respondent, and this is presumably not the point being made by Slembrouck. The second issue is whether the questionnaire allows respondents to comment on aspects of the programme that were of interest or concern to *them* (i.e. whether the right questions are asked). This is a point to which we return, below. And the third issue is whether the respondent feels able to tell the truth, or whether they are so constrained by what they think of as the norms of social behaviour that they desist from overt criticism or say less than they would like to. One factor which has a bearing on this last point is whether questionnaires are completed anonymously. Rowntree (1985), for example, claims that anonymous questionnaires permit respondents to express strong opinions that are unlikely to be expressed in public. Also relevant, however, is the relationship between the evaluator and respondent and that between respondent and evaluation, i.e. the extent to which respondents feel themselves to be stakeholders in the evaluation.

Saville and Andrews, who were LEA advisers, refer to the difficulty of preserving anonymity on small courses and the fact that after the course there would be continuing contact between themselves and course participants - factors which may have made the latter feel uneasy about being absolutely frank. Even so, Saville and Andrews felt it important to establish a means of communication with the teachers for whom they were responsible. They comment:

> They could ignore the form and contribute nothing; they could use it to express cynicism or sycophancy; or they could respond in kind to our openness and attempt to give honest, practical and helpful comment. What surprised us was how seriously course members generally took the task.
>
> (Rudduck 1981: 68)

Since the respondents in this case were local teachers who might go on to do other courses and can therefore be assumed to have had a vested interest in taking evaluation seriously this level of cooperation is not so surprising. (For counterevidence, however, see Appendix 7.2).

Let us return now to the criticism that questionnaires reflect the agenda of the institution or course leader rather than participants. This is most likely to be the case when a standard questionnaire is used for all programmes or all courses within a programme (a convenience but also a basis for the comparison of evaluation findings). Alternatives to the standard questionnaire that would answer this criticism include the posters and post-it slips discussed in section 2.6.3, the kinds of instrument included in Appendix 3.1 and questionnaires based on in-course soundings, formal and informal, a possibility that is explored in Chapter 7. All of these open-ended approaches allow individual voices to be heard. As part of a timetabled exercise on questionnaire design, participants might even be asked to contribute questions to 'their' own questionnaire. If one of the weaknesses of most questionnaires is that they only provide answers to the questions asked, then it would be preferable to allow participants to collaborate with other stakeholders in determining what those questions should be.

3.2.5 The lessons of experience

The experience of Saville and Andrews, reported in Rudduck (1981: 67), will probably strike a chord with most evaluators:

> at the beginning a questionnaire was all we had. At that time we considered it somewhat daring to ask participants to make comments, and the variety in our approach consisted entirely in the kind of questions asked, in the amount of blank space left under each heading, and in the style and timing of the distribution of the forms. The greatest drawback was that designing really effective response forms was beyond our competence. Another problem was how to interpret the data we were offered.

With time, and experimentation (see 3.2.3), they obviously became more proficient, but they also discovered - as do many institutional evaluators, perhaps - that they were increasingly able to predict participants' reactions. Time being at a premium, they then faced the dilemma of whether to continue to appear to allow participants to 'contribute to the shaping of future events' (*op. cit.*: 70), an opportunity which participants appear to have welcomed, or to deprive them not only of that opportunity but also of a potential model for the evaluation of their own courses. (They do not actually say how they resolved the dilemma, but we infer that they continued to use some form of questionnaire.)

One of their conclusions concerns the relative value of different kinds of data. The qualitative data elicited by open questions they found to be helpful in relation to future planning and as an indicator of differences in individual reactions and needs, but not as a reliable basis for immediate action since further checks would be needed on the generalisability of specific reactions. Nor do they set much store by quantitative data deriving from ratings and scales which, they say, 'provide little more than a rough-and-ready guide to the popularity of particular events or approaches' (*op. cit.*: 69).

Their general conclusion is also worthy of note:

> As a total evaluation of a course, questionnaires and rapid assessment forms do not seem to us to be all that effective. However, used to focus attention on particular aspects of the course or to find out how individuals are feeling at a particular moment, they may be valuable. In the jigsaw puzzle of understanding, they make a contribution that can *sometimes* be unexpectedly revealing about the things we are inclined to take for granted.
>
> (Rudduck 1981: 69-70, original emphasis)

As with any other uni-faceted approach to programme evaluation, questionnaire results need to be treated with caution. Tempting though it may be when participant comments are overwhelmingly positive to see these as *the* measure of the course, they are only one measure (with a variety of potential disadvantages) and at that, as Saville and Andrews observe, a measure only of some aspects of the course.

3.2.6 Conclusion

Despite some of their obvious advantages over interviews, questionnaires have had a rather bad press. Peil (1982: 97) cites Hopkins and Mitchell (1973) as saying that surveys are often questioned as a source of reliable, valid and/or theoretically meaningful information and some of the reasons for this have been indicated above. However, the main problem may be, as Peil notes, that they have tended to be conducted by those with little training or experience and have led to 'mediocre and misleading results which are passed off as indisputable fact' (*ibid*). If this is the case, it is more appropriate to level criticism at 'poor research design or poorly constructed instruments' (Berdie and Anderson 1974: viii) rather than the questionnaire *per se*. Given the frequency with which questionnaires are used in programme evaluation, it is clear that evaluators need an informed appreciation of the strengths and weaknesses of questionnaires in general and specific forms of questionnaire, and an awareness of how these can be combined with other data-collection instruments (see 3.1.3).

3.3 INTERVIEWS

> Interviewing is beguiling in its simplicity; anyone, it would seem, can ask a few questions to get someone else's point of view.
>
> (Powney and Watts 1987: 9)

3.3.1 Introduction

Like questionnaires, interviews are potentially useful at all stages in programme evaluation. Prior to a course, they are probably more frequently used for selection purposes than for evaluation; however, they can also be used within a baseline study (to establish attitudes, language level or methodological awareness, for example) and therefore contribute both to participant profiles and course design. Where it is not practicable to conduct interviews for such purposes prior to a course, they may take place in the first day or two. Such interviews can also serve additional purposes: where a choice of parallel courses is available, to establish whether participants are - or feel they are - on the most suitable course; to elicit first impressions and special interests or needs; or to modify participants' expectations when these are seen to be unrealistic. Once a course is under

way, interviews with individuals, even when primarily intended for the discussion of their performance or progress, can also yield insights into participants' reactions to assessment procedures and to programme content or processes.

The information obtained during individual interviews can, of course, be augmented by that from forms of oral interaction that differ as to scale (small group interview; plenary discussion), balance of interactants (staff-student meeting) or ostensible purpose (conversations over coffee; classroom simulations). Although the main emphasis in this section is on the individual interview, some reference is also made to group interviews.

3.3.2 Advantages and disadvantages of interviews

Table 3.2 sets out the major advantages and disadvantages of interviews.

Table 3.2: Features of interviews

	ADVANTAGES	DISADVANTAGES
economy		time-consuming, especially if transcription subsequently involved
resource requirements		unstructured interviews require special skills; recorder desirable
ease of administration		difficult to take field notes
reliability		only open to scrutiny if recorded; responses to open-ended questions pose problems of analysis
validity	completeness of response (*cf* questionnaire); probing can establish degree of consistency in responses	possible influence of interviewer bias and interviewer effect
comparability and generalisability		dependent on size and representativeness of sample and degree of standardisation of interview
utility	flexibility: interviewer can ask for amplification or explanation; well suited to elicitation of attitudes and opinions; linguistic problems can be overcome	tension between need for quantitative data (which implies degree of standardisation) and flexibility

As noted above, interviews can involve more than one interviewer and more than one respondent (3.3, below); they also vary along continua of structure and control (3.4, 3.5, below). However, for the purposes of general comparison between the interview and other

data collection instruments it will be assumed that the interview is a 'conversational encounter to a purpose' (Powney and Watts 1987: vii) between two people.

As we have seen in section 3.2.2, questionnaires can be completed under supervision or in the respondent's own time. One advantage of the interview over the latter type of questionnaire is that once the agreement of the respondent to an interview has been secured, there is a degree of control over the kind of data that is obtained, especially since interaction between the interviewer and respondent is possible concerning the form and content of the questions. The fact that individual questions can normally be followed up within the interview also means that topics can be explored in greater depth, a point of particular relevance if the intention is to elicit opinions or attitudes and the beliefs that underpin these.

The obvious drawbacks of the interview as a data-gathering technique are its cost, its openness to scrutiny (unless recordings are made) and its suspect reliability and validity, especially where - as would be common in small-scale INSET evaluation - the interviewer is involved in the evaluation design (self-fulfilling prophecies).

One factor in any interview may be interviewer bias, described by Powney and Watts (*op.cit.*) as the communication, explicit or otherwise, of the interviewer's own attitudes and behavioural characteristics which reflect these. (If we leave aside paralinguistic features, this may also be a problem in questionnaires, though one that is easier to detect.) The importance of interviewers preserving a neutral stance with regard to their own attitudes is emphasised by Henderson (1978: 89, citing Kahn and Cannell 1957):

> In his role as interviewer, he is not concerned with questions of agreement or disagreement with the respondent's sentiments; he is not concerned with the social and political implications of the information he is receiving; he is not concerned with moral issues which might be raised by the content of the interview. He is, for the period of the interview, amoral in this sense.

What might be termed 'interviewer effect' (respondents' perception of the differences between themselves and the interviewer) may also bias results, affecting the nature of the response (Langley 1987) as well as the degree of cooperation (Peil 1982). Among the personal characteristics of the interviewer that might have this effect, Powney and Watts

(*op.cit.*) list age, education, socio-economic status, race, religion, and sex. In an educational setting, such as INSET, role relationships may also play a part.

One possible consequence of the various influences listed above is that respondents are less than truthful. For instance, as has been noted in reference to questionnaires, they may give answers they think the interviewer wants (out of politeness or the desire to please). Peil (1982) lists a number of other possibilities. Respondents may:

ñ try to impress the interviewer
ñ try to save face by not giving answers that reflect unfavourably on them
ñ be reluctant to give truthful information for fear it may fall into the wrong hands (uncertainty regarding future use of data).

In interactions with teachers from other cultures, other totally unsuspected factors may be involved. Writing in an African context, Peil gives as another reason for untruthful answers the fear of danger: that listed children might be killed by spirits. It is not easy to imagine why in the course of an INSET-related interview one might wish to ask an African teacher to list his or her children, but some understanding of cultural *mores* is clearly advantageous.

Powney and Watts explain the mechanisms generally at work in this way:

> To some extent, all interviews are seen as threatening by those being interviewed. The person being interviewed makes some kind of judgement about the interviewer and the kind of definition of themselves and their situation that they want to project. It is a decision as to **which layer of truth** they will make accessible.
>
> (Powney and Watts 1987: 44-45, emphasis added)

Henderson (1978: 80) refers to Dean and Whyte (1958), whose view is a little less jaundiced:

> any statement represents the individual's perception of reality, moderated by his cognitive and affective reactions to his situation and reported through the medium of his personal use of language.

For an evaluator, it is important to distinguish between subjective data and objective-like data. Subjective data is information on an informant's emotional state, opinions, attitudes, values, behavioural tendencies, any of which may be inconsistent over time or even within a single discourse. Recognised for what it is, inconsistency is not necessarily a problem. According to Henderson, 'the true or "valid" picture is this teacher's total position, complete with all its apparent logical inconsistencies' (Henderson, *op. cit.*: 82). There is, however, a problem with subjective data which takes the form of a retrospective account in that, as Henderson acknowledges, 'there is a common tendency to modify the recollection of past experiences to make it more congruent with a current point of view' (*ibid*). Unfortunately, objective-like data may be no more reliable. The retention of observable 'facts' is also subject to such frailties as loss of concentration, memory loss or blurring or even prejudice, and the informant may concoct an account that seems to fit the bill (Henderson, *op.cit.*). The evaluator, Henderson concludes, needs to be aware of such possibilities, assess the plausibility of what is offered, and if possible incorporate cross-checking mechanisms.

3.3.3 Administration

One practical consideration in decisions concerning interviews is the number of people who will be involved. Although one-to-one interviews are probably the most common, group interviews (or discussions) are used, and other variations are possible (e.g. one interviewer to two informants or two interviewers to a single informant).

Each configuration has advantages and disadvantages, as Powney and Watts (*op.cit.*) point out. One-to-one interviews are easier to manage; issues which arise can probably be kept confidential; and because there is only one set of responses to deal with, analysis is more straightforward (the views of the individual are clearcut, and there are no interpersonal influences). Group interviews, which have the advantage that they allow a range of responses to be collected economically, are most suitable when the group has been together for some time, since individuals will feel more at ease with each other. Audiorecording the discussion, despite the initial inhibitions that some might feel, is almost certainly the best way of obtaining a reliable record of what is said. However, one practical difficulty is that it may be difficult retrospectively to attribute comments to individuals. Note-taking and videorecording are a possibility, of course; an alternative is to

have a second interviewer present who simply takes notes or who alternates this role with the first interviewer.

Henderson (1978: 86-87) sees two further advantages of the group interview beyond the obvious one of collecting 'a wider range of experiences and responses' than the individual interview:

1. 'group interaction may serve to remind individuals within the group of details of experience which may otherwise have been forgotten';
2. a skilful interviewer may be able to induce individuals to make progressively open and personalised comments.

He also notes a number of potential disadvantages:

1. the public setting may inhibit some participants;
2. there is a risk of domination by the few, who may also influence the rest to express similar views;
3. discussion may be tangential and superficial.

A number of specific techniques have been devised to overcome the second and third problems listed by Henderson, i.e. for the purpose of structured discussion and decision-making within group settings. These include Delphi Decision Making, the Goldfish Bowl, and the Nominal Group Technique, of which the latter is probably the most appropriate for the purposes of programme evaluation. As described by Newstrom and Scammell (1980), the Nominal Group Technique has the following stages:

1. those present are divided into groups of 5-6;
2. an open-ended task is set;
3. individuals brainstorm silently and jot down their own ideas;
4. the ideas are recorded on a flipchart, one idea per person at a time (clarification is encouraged, no criticism allowed);
5. individuals evaluate and vote for ideas (5 points for the best idea, 4 for the next best, and so on);

6. individual votes are tabulated and a group report is prepared giving the conclusions;
7. the report is presented to the other groups.

Delbecq and VandeVen (1971) are acknowledged as the source of this idea.

Newstrom and Scammell point out some advantages and disadvantages of the technique: on the one hand, it 'allows voting anonymity; provides opportunity for equal participation of members; eliminates distractions of other group methods'; on the other, 'opinions may not converge in the voting process; cross-fertilization of ideas is constrained; the process may appear mechanical' (*op.cit.*: 107). For a course tutor or organiser seeking feedback on a course, disadvantages such as lack of discussion (the explanation for the name given to the technique, perhaps) might seem less significant. What would emerge from such an activity is a clear picture not only of the range of points considered relevant by individuals but also the extent to which these appear salient to the whole cohort. Data of this kind might be forthcoming from a questionnaire, but the key difference is that the points evaluated in the Nominal Group Technique are those that course participants have themselves raised.

3.3.4 Categorising interviews

Another dimension to the interview is the extent to which it follows a predetermined pattern. At one extreme is the 'structured' or 'formal' interview, in which the content and procedures have been established in advance and the respondent's answers are typically entered on a schedule (Cohen and Manion 1994). At the other end of the continuum, for Cohen and Manion, is the 'non-directive' interview, which is based on the therapeutic or psychiatric interview (i.e. the interviewer's purpose is to induce the respondent to present his/her reality without any external structuring). Between the two lie two more categories. One is the 'unstructured' interview, which is unstructured only to the extent that the interviewer feels free to phrase questions in whatever way seems appropriate, to modify the sequence of topics to be touched on, to follow up specific answers, and in general respond in a flexible manner to the evolving interaction. The other is the 'focussed' interview. This is described by Cohen and Manion as being a more controlled form of the non-directive interview. The interviewer has certain hypotheses concerning the experience undergone by the respondent and focuses on their subjective responses to that experience.

Other writers, such as Henderson (1978), appear not to distinguish between less structured and focussed interviews. Concerning the possibly more real distinction between the highly structured and the less structured interview, Henderson expresses the commonsense view that in the context of INSET evaluation the highly structured interview is likely to be less useful since it 'fails to capitalise on the principal advantage of the interview over the questionnaire, that the former is much more flexible and therefore potentially a much richer source of data' (Henderson 1978: 86).

However, relatively unstructured interviews, where the interviewer has decided only a broad line of questioning in advance, are in Henderson's view, riskier than structured interviews. He cites Hyman *et al* (1954), who claim that results may be biased by two kinds of interviewer expectation. An informant's answers at an early stage in an interview can lead the interviewer to interpret subsequent responses in a particular light ('attitude-structure expectation'); alternatively, the impression the interviewer gains of the kind of person the informant is may lead to 'role expectation' ('interviewer effect' in reverse). The consequence of either type of bias, Henderson points out, may be selective listening or even selective note-taking. An audiorecording would compensate for any deficiencies in these respects, but could not make up for any deficiencies resulting from the wrong questions being asked. These strictures apply not only to a single interview but also to the danger of generalising to a series of interviews expectations based on a single experience.

3.3.5 Structure *vs* control

Powney and Watts (1987) suggest that such distinctions as structured/unstructured and focussed/unfocussed (together with limited/in-depth), are less significant than those concerned with who controls the interview, the issue raised by Slembrouck in relation to the questionnaire. They therefore propose a distinction between (1) 'respondent interviews' and (2) 'informant interviews'. In (1) the interaction is structured by the interviewer; it will revolve around discussion of issues that concern the interviewer and typically include specific questions to be answered. The interviewer sets the agenda. In (2), on the other hand, the interviewer may deliberately relinquish control and simply adopt a 'let's wait and see' attitude, the purpose being to discover what is uppermost in the mind of the informant. The source of approach 1 can be found in Booth's survey research in the

1880s; approach 2 derives from the clinical interview typified by Piaget and Freud, and ethnomethodology, which itself owes much to anthropology (Powney and Watts 1987: 19).

3.3.6 Conclusion

If questionnaires are the most appropriate means of obtaining certain types of (typically quantitative) self-report data from a large population, then interviews are the most obvious way of obtaining qualitative self-report data from a small population. Some of the possible drawbacks of interviews have been discussed in section 3.3.2; their strength is that, carefully planned and skilfully executed, they can give access to data that would be very difficult to obtain by any other means.

As is the case in any research endeavour, the practice of interviewing, in the broad sense adopted by Powney and Watts (1987), is a constant attempt to maximise the potential of the procedure while guarding against the more obvious risks. As Powney and Watts reflect:

> ... it is always a salutary exercise to re-examine one's own personal performance, note disparities, indiscretions and rank failures, and to develop remedies. Interviewing, as a systematic art, needs constant practice and appraisal.
>
> (Powney and Watts 1987: Preface viii)

3.4 PARTICIPANT JOURNALS

3.4.1 Introduction

The use of diaries (or journals) as a research tool, to investigate learners' preoccupations, learning strategies and reactions to instructional processes, is a relatively recent technique in second language education (Bailey 1983). The first phase in this research was characterised by a focus on diaries kept by researchers of their own (language) learning processes; this 'diarist as researcher' stage (Palmer, C. 1992: 228) was followed by the 'diarist as participant' stage (Palmer, *ibid.*), in which the diarist was a course participant writing for a tutor. In one variation on the latter, 'the dialogue journal' (Brinton, Holten and Goodwin 1993), the tutor engages in a form of correspondence with the diarist through the medium of the journal.

A number of different terms are used in the literature to refer to data of this kind. 'Diary' and 'journal' are used more or less interchangeably, as a matter of style or personal preference. However, Jarvis (1992) points out that the connotations of such terms with personal, private writing may lead to a reluctance to share that writing with a course tutor; for this reason, she abandoned her first choice of term, 'learning diary', for 'learning' record'. This did not remove the tension, she adds, but it may have lessened it. Other writers, for similar reasons, refer to 'logs' (Thornbury 1991) or 'learning log' (Porter, Goldstein, Leatherman and Conrad 1990); the latter state explicitly that 'the journal is not a personal diary' (*op.cit.*: 229). In an attempt to avoid any further confusion, the term 'journal' will be used in the remainder of this section.

The emphasis in most reports of journal use in teacher education programmes has been on the journal as a pedagogic device, a means by which participants learn through reflecting on their own teaching (Haill 1990; Thornbury, *op.cit.*), the relevance of the course to their own teaching (Porter *et al*, *op.cit.*; Palmer, G. 1992) and/or course content and processes (Palmer, C., *op.cit.*). Very few studies have looked specifically at participant diaries as input to programme evaluation. The exceptions are discussed below.

3.4.2 Journals and programme evaluation

Murphy-O'Dwyer (1985) seems to have been the first to propose that journals might serve as an instrument for the evaluation of teacher education programmes. She describes a carefully organised study in which daily diaries were kept by fifteen teachers on a two-week INSET course, by tutors on the course, by an observer and by the researcher, also acting as observer. As presented, the results of all this effort are somewhat disappointing: the preoccupations of the participants closely resemble those of previous diarists, many of whom have also been teachers, and little light is shed on the journal as a form of data collection with particular relevance to programme evaluation.

Hundleby and Breet (1988) describe their use of 'methodology notebooks' on a one-year INSET course in China in which only two hours per week were allocated to methodology and the notebooks were conceived as a way of extending the discussion of pedagogy beyond the limits of the classroom. One of the purposes for which the notebooks were

exploited was to identify areas in which individual students were interested as a basis for the development of individualised reading programmes. Hundleby and Breet comment: 'This was particularly valuable for students whose future teaching needs differed considerably from those of the majority' (35). Over time, it was also possible to observe how trainees' attitudes changed. Thus, initially negative reactions to pairwork were transformed into a positive awareness of the increased practice opportunities generated. Although the term 'evaluation' appears nowhere in their paper, Hundleby and Breet were clearly sensitive to the kinds of evaluative information they were collecting and were using this for course development purposes and as a measure of the effectiveness of the course.

The basic rationale for the 'learning records' kept by participants on the three- and four-month UK INSET courses run by Jarvis (*op. cit.*) was that reflection on learning experiences would lead to increased reflection-in-action. Following a helpful exemplification of some of the difficulties involved in getting participants to reflect critically in writing, a competence which some never achieved, she uses quotations from participant evaluations to indicate the value some attached to these records. Her own assessment of their value is based on two sources. The records themselves gave her the feeling of 'being in contact with my learners' learning' (142): their attitudes to the course, their anxieties, and - since these were learning records - what they felt they had learned. Her observations of former participants working on in-country seminars and the reports she has received of them from others lead her to suggest that those who used the learning record for reflection were subsequently able to make positive changes in their practice. About this second form of evidence she is more tentative, but in that it relates to longer-term effects and does not rely on participant self-report it is possibly more convincing.

Two related articles by Christopher Palmer (1992) and Gillian Palmer (1992) are more specifically related to the potential of journals for programme evaluation purposes. Christopher Palmer points out that while the journal-keeping reported in previous studies has

> provided many useful insights into how a programme is received, the quality and quantity of the evaluative comments obtained is not always satisfactory, being on occasion superficial and inconclusive and hence effectively limiting the scope of the evaluation.
>
> (Palmer, C. 1992: 228)

He therefore argues that participants be given clear guidelines for the content of the journals and is logically consistent in providing a carefully detailed record of the procedures used in his own study, an investigation of the use and feasibility of the journal as a tool in self-assessment and programme evaluation. One feature that is of particular interest in the context of the present chapter is the use made in Palmer's project of questionnaires in combination with journals (see Figure 3.1, below):

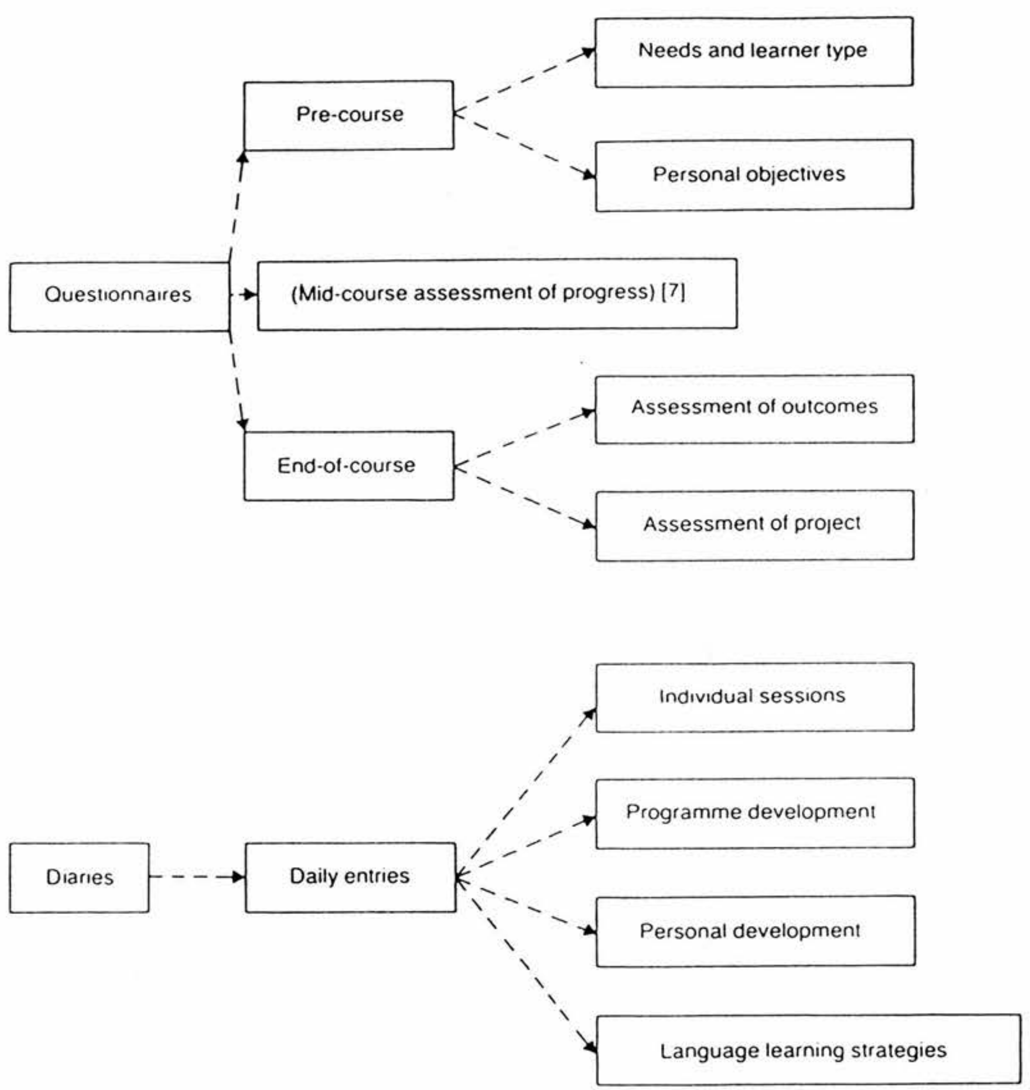


Figure 3.1: Diagram showing the organisation and evaluative/pedagogic objectives of the study

Questionnaires were used to draw up an initial profile of participants and their objectives; these were subsequently used as a yardstick for measurement of progress during and at the

end of the course. Palmer's conclusion is that the journals fulfilled 'a very important role in helping the course organisers to assess the success of the programme as a whole *in a way not previously possible* (*op.cit.*: 235, emphasis added). Of even more significance, perhaps, is the following end-note:

> The mid-course questionnaire was used on only one teacher's course as it was found to be unnecessary with assessment of progress being adequately covered within the diaries and a mid-course questionnaire running the risk of overkill!
>
> (Palmer, C. 1992: 235, note 7)

Evaluation makes demands on participants and tutors. As Palmer implies, the deployment of multiple instruments can therefore only be justified if these result in complementary data.

Writing of the same context, short courses in the UK for Norwegian teachers, Gillian Palmer's stated focus is 'the practical feasibility of the journal as a pedagogical and evaluative tool' (240). Evaluation of this programme had traditionally taken the form of an end-of-course questionnaire, a 'forum' on the final day, and individual reports written by participants for the Ministry of Education on their return home. However, the need was felt for some means of monitoring participant reactions during courses, and journals were surmised to be a suitable instrument for obtaining these, as well as fulfilling pedagogic purposes. Palmer explains what was expected of participants and the intended outcomes:

> Participants were asked to evaluate the teaching sessions, taking into consideration how far the teaching matched their own individual expectations and needs and its relevance to their own teaching situation. Monitoring of such feedback would allow for changes to be implemented during the course itself, as well as providing suggestions for future courses.
>
> (Palmer, G. 1992: 242)

Participants were given a choice between keeping individual or joint journals; course tutors also kept journals, and the researcher kept a record relating to the pilot project.

Her findings concerning the feasibility of including a journal as a course component are somewhat negative, time and pressure being particular problems for both participants and the researcher. Nevertheless, the journals did appear to fulfil the intended pedagogic and evaluative functions: Palmer claims that they stimulated participants to relate course input

to their own teaching and encouraged a 'conscientious approach to course evaluation' (247). Indeed, participants 'expressed satisfaction at being able to comment daily on the course in as much detail as they wished' (*ibid.*). The following classification is offered (*op.cit.*: 249) of participants' evaluative comments:

1. Suggestions which could be immediately implemented.
2. Suggestions of relevance for future courses.
3. Suggestions which provided otherwise hidden perceptions and insights of use to tutors.

Palmer concludes that the journals were 'a successful innovation' (249).

3.4.3 Advantages and disadvantages of participant journals

The major advantages and disadvantages of participant journals are shown below, in Table 3.3:

Table 3.3: Features of participant journals

	ADVANTAGES	DISADVANTAGES
economy		data analysis time-consuming
resource requirements	no special resources required	
ease of administration		where evaluator is to respond but is also teaching, pressure of time
reliability		variety makes analysis difficult
validity		time pressure may mean task is 'tokenized'; participant suspicion regarding evaluator's intentions may mean entries are saccharine; recall may be suspect
comparability and generalisability		comments likely to be very disparate; may be difficult to quantify
utility	allow for continuous monitoring of individual reactions and learning on the basis of which informed changes can be made, help given and learning for participants maximised; may also indicate general issues and themes	respondents may be reluctant to evaluate

Most of the sources cited above draw attention to the cognitive value of journal-keeping. Porter *et al* see it as serving both a cognitive function - 'writing both stimulates and shapes ideas' (*op. cit.*: 234) and a 'social' function; for Brinton *et al*, the latter is more a matter of establishing a relationship between the (participant) writer and the (tutor) reader. These and the other sources referred to above also contain a number of scattered references to the possible contribution of journals to programme evaluation. Apart from the points included in Table 3.3, these can be summarised as follows:

Journals can:

1. be particularly revealing in the case of course members who are reluctant to participate in class (Murphy O'Dwyer 1985) or to express doubts openly (Haill 1990);

2. reveal what goes on in participants' own classrooms or homes (Haill, *op.cit.*);

3. encourage self-evaluation (Bailey and Ochsner 1983) and therefore promote more objective programme evaluation (Palmer, C., *op.cit.*);

4. permit triangulation of source (viewpoints of, e.g. participants, tutors, observer) (Bailey and Ochsner, *op.cit.*; Palmer, C., *op.cit.*; Palmer, G., *op.cit.*) and method;

5. where writers are teaching while following a course, can provide feedback on the application of ideas presented in the course (Haill, *op.cit.*);

6. allow participants to feel involved in programme evaluation on a continuous basis (Palmer, G., *op.cit.*).

As is the case with other data collection procedures, questions can be raised concerning the reliability of participant responses. Where the writer is writing for a known reader, 'fulsome praise' (Haill, *op.cit.*: 9) is a possibility, but Haill claims she overcame this potential problem by giving very specific prompts (e.g. not 'Was the class useful?' but 'What was useful?' or 'Why was it useful?'). Christopher Palmer (*op.cit.*), investigating the effect of anonymity of writer and the effect on entries of the relationship between writer and reader (tutor *vs* researcher) found that writers were much more concerned about

anonymity when they were writing for an unknown researcher; when they were writing for their own tutor they made no effort to conceal their identities. Gillian Palmer (*op.cit.*) asked participants in an end-of-course questionnaire whether anonymity helped them to express their feelings more freely. 55.5% of those who had written for an outside researcher claimed that anonymity had helped, whereas 61.5% of those who had written for an inside researcher stated that anonymity would have helped. Flying somewhat in the face of these figures, Palmer concludes that there is little difference between the two groups, and that 'the security and trust which a known figure provides ... appear to outweigh any difficulties in ... evaluating a course on which the reader is also a tutor' (*op.cit.*: 250). She nevertheless recognises that the outside researcher may have an important role to play as an impartial assessor.

Given the general enthusiasm for journals evidenced in the literature, it is reassuring to find an occasional cautionary note. There may be participants who are openly hostile to the notion of journal-keeping (and who, it is generally agreed, should not be compelled to take part), and there may be others who seize the opportunity for critical comment only too readily:

> the diary is a sensitive tool and the engendering of a positive critical attitude through the diary is also a reflection of the atmosphere created on a course. It is highly possible that the reverse would occur if a diary project is not handled carefully.
>
> (Palmer, G. 1992: 248)

3.4.4 Conclusion

Journals have been justified on the grounds that they benefit the participant; their potential as programme evaluation instruments has been less frequently referred to. From a course tutor's point of view, journals make possible a greater sensitivity to participant concerns and needs and an immediate response through the personalised medium of the journal or through more general action. One problem is that if the dialogue journal format is adopted on a short course, this may place what is perceived as unreasonable pressure on participants as well as proving a burden for tutors. Any economies result in loss, but if the journal is seen primarily as a personal record, rather than one side of an exchange, and the pedagogic purpose clearly distinguished from evaluation, this may take some of the pressure off both participants and tutors and a reduced evaluative purpose may still be

served. One way in which the this can be achieved is illustrated by Buckley and Caple's (1990) 'course log' (Appendix 3.3) . The log format first requires participants to formulate their expectations of the course and keep a daily record of what they have learned; at the end of the course they refer back to their expectations, read through their daily comments and write an evaluation indicating what they feel they have learned from the course. The evaluation is then handed in, possibly after discussion with other participants. From an evaluation perspective, the drawbacks here, of course, are that because tutors have no access to participants' logs, they are unable to use these as a means of monitoring participant learning while the course is in progress, and the final comments may be too general or too individual to be of any help as far as programme development is concerned.

The dialogue journal has the further potential disadvantage that it might be used by participants for communication on personal as well as pedagogic issues, encouraging emotional dependence (Rinvolucri 1983, writing of pupils' diaries; Haill 1990). This is perhaps less likely with mature participants. When Gillian Palmer notes that the journals allowed tutors 'to deal with ... problems before they took on significant proportions' (*op.cit.*: 248), she appears to be referring to *programme*-related problems, and one of her appendices lists changes made to the course as a result of journal comments.

3.5 PARTICIPANT PLANS

Questionnaires are the most obvious way of collecting structured written feedback from participants. Journals can either elicit naturalistic data or be used for more directed data-collection. This section briefly considers a number of other techniques for eliciting data from programme participants, all but one of which involve the participant in formulating a plan for post-course action.

3.5.1 Letters

Newstrom and Scammell (1980), writing in the context of industrial training, suggest that three types of letter may be useful in facilitating transfer of training (i.e. the application of knowledge and skills acquired on a course to the work context). At least two of these also lend themselves to programme evaluation.

3.5.1.1 Letter to the trainer

Apart from the aim referred to above by Newstrom and Scammell, this type of letter is intended to 'feed back information on the concepts found to be most useful on the job' (*op.cit.*: 293). The 'letter' (which may take the form of a blank sheet of paper, journal or structured response sheet - see Appendix 3.4) is given to participants at the end of a course with the request that it be returned by a specific date (e.g. in a month's time). From an evaluation perspective, the letter therefore functions as a post-course questionnaire. Newstrom and Scammell outline the *pros* and *cons* of this procedure:

> The danger is the low response rate that is likely unless a follow-up is used, or participants are highly committed. The survey format permits the use of greater structure (*sic*) to the responses, and is more likely to obtain a respectable response rate and meaningful data if it is kept brief. The journal approach requires considerable persistence on the part of the participant. However, if rigorously followed the journal may contain some of the most useful spontaneous insights. The main point, of course, is that the trainees, having been alerted to the certainty of *some* follow-up mechanism, will be more likely to retain their newly-acquired knowledge and practice their new skills.
>
> (Newstrom and Scammell 1980: 293, original emphasis)

Vance (1979) is acknowledged as the source of this idea.

3.5.1.2 Letter to the boss

Whereas the 'letter to the trainer' is a backward-looking device, that 'to the boss' looks forward. Its purpose is to 'capture' and 'channel' the enthusiasm generated during a course 'toward the improvement of on-the-job performance' (Newstrom and Scammell, *op.cit.*: 297). Participants are given the pro-forma (see Appendix 3.4) at the end of the course and asked to complete it before they leave. This 'letter' is then given to their superior at work as a basis for discussion. Two 'discussion questions' are suggested (*ibid.*):

1. What factors will serve to prevent you from implementing the desired changes ? (e.g. non-supportive supervisor, time pressures, irrelevant material)

2. What steps can you take to ensure the likelihood of changing your behavior ? (e.g. develop a support group or buddy system, solicit your supervisor's support, attend follow-up session).

It is not clear when these questions would be used; one can only assume that they are intended to be considered after the form has been filled in, but the indication of time required (5-15 minutes) hardly allows for this.

No comment is made in this case on the advantages or limitations of the technique. From a pedagogic perspective, the most obvious strengths would seem to lie in its value as a focussing device and as a spur to action. The major disadvantage might be the severity of the constraining factors and the apparent failure to highlight those factors over which the participant can exercise some control. From an evaluative perspective, the technique would work well if it were combined with a pre-course analysis of wishes and expectations, since it would be possible to discern the extent to which individual participants felt these had been fulfilled. Follow up would obviously be desirable, however, to gauge how far good intentions had been translated into short-term action (the delivery of the letter to the boss and discussion of its contents) and longer-term action in the form of behavioural change related to the course.

3.5.1.3 Letter to myself

This is intended to serve the same broad pedagogic purpose as the first two letters. Where it differs principally is in the attempt to engage participants in a form of contract with themselves to achieve particular (behavioural) post-course goals. Following discussion of the need for change and the difficulty of bringing this about, participants write a letter to themselves, seal it, and hand it to the course organiser, who undertakes to send it to them in 30 days' time.

The disadvantages of the procedure, as far as programme evaluation is concerned, will be clear: the course organiser does not see what participants have written and unless an attempt is made subsequently to discover both what their goal was and what progress they have made in attaining it, has no feedback on the course itself. Even if such information could be collected, it is possible that the goals formulated will have little direct relevance to course content.

From a course provider's point of view, a more feasible variation on this idea is the kind of action plan which is described below.

3.5.2 Action plans

Like participant journals, action plans - what one intends to do as a result of what one has experienced - can be used to encourage and focus reflection on a course that is being followed while participants are teaching (Estaire 1993). The potential also exists, however, for using the action plans formulated by participants during or at the end of a course as the basis for post-course evaluation by programme tutors. Alderson (1985a), taking as an example a 10-week course for teachers of ESP, describes a three-stage process (formulation of plan by participant - discussion with tutor - follow-up letter) designed to provide information on the extent to which participants were able to implement what they had learned while in the UK. An important feature of the planning stage was that participants were asked to consider the potential difficulties that might be involved in implementation and how they would attempt to overcome these. (Alderson's example of the briefing notes given to participants on this course is included as Appendix 3.5.)

In the field of industrial training (= teaching), action plans appear to be better established. Sheal (1989), for example, suggests that action plans might be used in end-of-course evaluation as a focus for discussion and as a basis for follow-up. He recommends that the plan should include 'the objective of each action, the timing, the names of the people involved, the resources required, potential obstacles and solutions, etc' (175). This is essentially a refined version of the implementation plan described by Alderson (*op.cit.*), but without the negotiation, perhaps because industrial training courses are normally of very limited duration. Buckley and Caple (1990) point out in passing that the action plan can serve a pedagogic purpose in facilitating transfer of training. Like Sheal (*op cit.*), however, they also see its potential in end-of-course and post-course evaluation. At the end of the course, the plans, which in their view should consist of a series of prioritised statements, are scrutinised with a self-critical eye by the instructor or instructional team:

> if the trainees' action plans did not include most, if not all, of the important learning objectives, then serious questions would have to be asked about the adequacy of the preceding needs analysis and any other analyses that may have been undertaken.
>
> (Buckley and Caple 1990: 194)

This is, we should remember, a training situation, and one in which there appears to be some assumption of common needs and starting points. Subsequently, the plans will be 'an

invaluable aid to subsequent job and performance related evaluation of training' and serve in particular as 'the basis for the design of the follow-up questionnaire or interview schedule' (*ibid.*). A record should be kept, they say, of how much of the plan was implemented, which skills were involved, and what was 'sidelined' and why.

3.5.3 Advantages and disadvantages of participant plans

The major advantages and disadvantages of participant plans are presented in Table 3.4, below:

Table 3.4: Features of participant plans

	ADVANTAGES	**DISADVANTAGES**
economy		time may be needed during course for negotiation
resource requirements	no special resources required	
ease of administration		time required post-course may conflict with other commitments
reliability		reported action unverifiable
validity		affected by return-rate; response may be affected by role relationship
comparability and generalisability		major problem of comparability even if standard format adopted since individual circumstances (and perhaps plans) may vary greatly
utility	insight into what is valued, what gets implemented and (potentially) where things do not, why not	

Despite some of the obvious disadvantages associated with self-report methods and post-course evaluation and the likelihood of very diverse responses - especially on longer courses for teachers from very different teaching contexts, this method shares an important positive feature with unstructured interviews and participant journals: it elicits what participants feel to be important, and it does so at a time when they are in a position to make informed summative judgements (at the end of a course, when the whole scheme of things is clear) and, if there is a follow-up phase, when there has been an opportunity for implementation. With the exception of McGrath (1996), the L2 teacher education literature contains no documented accounts of the use of participant action plans in programme evaluation. Their potential is explored in Chapter 8.

3.6 OBSERVATION

> what people say is notoriously different from what they do
>
> (Alderson and Scott 1992: 54)

3.6.1 Introduction

Observation fulfils different purposes at different stages of the evaluation cycle. If the objective is to evaluate effects of particular programme conditions and/or processes, the evaluator may observe prospective course participants teaching a short time before the course begins in order to collect baseline data on individual and collective practices. This can then be compared with observation data collected during and after the course. Such a sequence is better suited to a course in which the emphasis is on training; where course aims are more broadly educational, the effects may be less tangible as well as less immediate. Apart from this more obvious focus on participants and observable learning outcomes as an indicator of programme effectiveness, in-course observation by the Course Director or someone external to the course may also contribute to the evaluation, formative or summative, of programme personnel and course processes.

As a means of assessing outcomes in terms of participant teaching behaviours, during a course or following a course, observation has one great advantage over questionnaires and interviews: it reveals what teachers do rather than what they think they do or say they do. Used post-course and at a suitable interval or intervals, it also reveals the degree of permanence of changes seen during a course (Henderson 1978). This is not to argue for observation to the exclusion of other data collection instruments - in fact, to combine interview with observation could result in an intriguing picture of the complex interaction between 'espoused theory' and 'theory-in-use' (Argyris and Schön 1974). It is, however, an argument for observation as a more reliable source of information on teaching behaviour than self-report.

3.6.2 Observation as a means of evaluating implementation

Programmes are implemented by programme personnel; the ideas or behaviours acquired during these programmes are then implemented (or not) by course participants. Beretta

(1992: 12) is thinking of the first type of implementation when he writes: 'The way that programs are implemented is fundamental to evaluation. The most obvious way of getting this information is through observation'. Alderson and Scott (1992), writing of an ESP project in Brazil, make a similar point, if less strongly:

> The absence of direct classroom observation data is problematic, since we are forced to rely upon inferences from reports of behaviour .... Since what people say is notoriously different from what they do, some observational data would have been very valuable, if only as corroboration.
>
> (Alderson and Scott 1992: 54)

Although both quotations concern language programmes they also have relevance for INSET programmes. Beretta's unequivocal statement follows a discussion of the much-criticised product-oriented methods studies of the 1960s, in which claims were made about the superiority of one method over another without any attempt to establish the degree of correspondence between 'ideal' and actual classroom practice. The point here, and in the quotation from Alderson and Scott, is that to make informed evaluative statements about certain aspects of a course we need to have a basis of objective fact, in so far as this is obtainable. The use of observation to obtain objective data is discussed under 3.6.4, below. We turn now to the second type of implementation and another purpose for observation.

The point has already been made in Chapter 2 that in general the summative evaluation of INSET programmes is most appropriately carried out some time after a specific course has been completed. It seems indisputable that where the focus of the course has been on behavioural change (e.g. changes in teaching practices) the most appropriate means of evaluation will be observation of participants teaching in their own classrooms. Since the observer is likely to have a clear sense of what he or she expects to see (evidence of take-up from the course) a structured observation checklist will normally be used. Such a checklist is also indispensable in order to ensure standardised data collection and cross-participant comparison.

The checklist should encapsulate what Weir and Roberts term the 'criterial features' of the programme, defined as 'the key defining characteristics of a programme as planned' (Weir and Roberts, *op.cit.*: 176). These, they say, can be identified by such means as prior

unstructured, exploratory observation; records of course plans and current implementation (in meaning 1, above); and discussion with insiders (to get at the 'informal orders' referred to by Holliday (1992)). Some of the issues concerning the design and use of observation instruments are picked up below in 3.6.4.

3.6.3 Participant observation

In some cases, the focus of evaluation activity may not be the effect of the programme on individual participants but the programme itself, its rationale, aims, organisation and processes. There are obvious advantages in using for this purpose an external evaluator who functions not simply *post hoc* but as a participant (or non-participant) observer.

The term 'participant observation' was coined by Lindeman (1924), a sociologist (Henderson 1978). In participant observation, 'the researcher becomes a member of a group or institution and records what it is like to be actively involved in the events which he is studying. By describing "from the inside", he has access to information which would probably be denied to an outsider' (Nisbet and Watt 1984: 75). In the context of programme evaluation, the evaluator is briefed in advance by the course organiser and then acts - overtly or covertly - as a course member, participating fully in the course and interacting naturally with other course members.

It is perhaps easier to justify the use of covert methods in sociology than in educational evaluation and, as Nisbet and Watt point out, there may be difficulty in reconciling the role of active participant 'with the detached neutral role of the conventional researcher' (*ibid*). An active participant observer may be able to monitor his or her own reactions subsequently but it is dubious whether they are equally capable of attending to other participants. It is likely, moreover, that informal conversations with other participants will shade into interviews or be perceived as such. These potential problems are not acknowledged in Henderson's evaluation of a course in pastoral care (reported in Henderson 1978); however, there seems little doubt that an observer-evaluator will find it more convenient to be non-participant if there can be some degree of certainty that this will not affect the participation of others. This point is taken up in Chapter 7.

In relation to interviewing participants, Henderson (*op.cit.*) sees the participant observer as having certain advantages over someone who has not been present at the course. Such an observer (who could presumably also be present but non-participant in the full sense of the term) can:

- pick up features of the interaction of which participants are unaware because they were 'unintended' or 'unrecognised' (*op.cit.*: 90);
- ask more focussed questions and 'generally get at depth material more successfully' (*op.cit.*: 91);
- select informants on the basis of what is known of their skills and insights;
- 'absorb information that may seem irrelevant at the time but may turn out to be valuable during subsequent analysis';
- 'reformulate the problem as he goes along' (*ibid*).

'Other methods', comments Henderson in conclusion, 'rarely approach this adaptability to the study of social dynamics' (*ibid.*).

The main disadvantage of participant observation is that, drawing as it normally does on such a rich set of data sources, it is not easily accessible to public scrutiny. There is also an issue of observer bias. Henderson (*op.cit.*) surveys a number of examples of evaluative reports written by teacher participants, by tutor participants and, in one case, a pair of reports by a tutor and two participants on the same course. In respect of differences between the latter, he raises the issue of objective truth and informant bias (see 3.6.4, below, and discussion under 3.3.2 of subjective reality).

3.6.4 Advantages and disadvantages of observation

The major advantages and disadvantages of observation are shown below, in Table 3.5:

Table 3.5: Features of observation

	ADVANTAGES	DISADVANTAGES
economy		very time-consuming, even if on-site, especially if checklist needs to be designed and piloted and there is need to view, transcribe and analyse recordings; off-site, much time

		may be spent travelling
resource requirements		checklist design requires special skills, and users training; recording equipment desirable and possibly technician if video used.
ease of administration		off-site observation subject to constraints frequently beyond control of evaluator; in participant observation, difficult for observer to participate fully and take field notes
reliability	recording (if possible) represents hard evidence	one-off observation may be unrepresentative; richness of data a problem for observer without checklist; when checklist used, accuracy of analysis subject to limitations of instrument and skills of observer; difficult for participant observer to fulfil demands of both roles
validity	record of what happens (*cf* self-report)	risk of observer bias, even when recording available, and observer effect
comparability and generalisability		comparison across cases difficult, even if standard checklist used, since contexts very different; logistical factors may limit size and representativeness of sample
utility	observer sees/recordings reveal things of which interactants are unaware	

From the evaluator's point of view, the main advantage of observing individual participants in their own work contexts is that he or she is dealing with reality (of a sort) rather than reality at a remove (self-report, course documentation). As indicated in Table 3.5, however, the potential disadvantages are quite numerous. Apart from cost and convenience, there are associated problems of reliability and validity.

Writing of observation for assessment, Wallace (1991) sees reliability as 'an acute problem' and suggests that as an absolute minimum participants should be observed 'on at least two occasions at different levels, and ideally by different assessors' (1991: 130). Weir and Roberts (*op.cit.*), referring to the use of structured observation instruments for purposes of research and evaluation, distinguish between intra-observer reliability (same observer, same event - videorecorded, and therefore observable on different occasions) and inter-observer reliability. They comment that even if consistency is achieved in these

respects validity remains a problem. Problems of validity may be associated with the instrument, which is invalid if it fails to capture certain relevant data, or the observer. All observation is selective and where the observer is not using a structured checklist, as in an in-depth study of a course in process, observer bias may result in certain relevant data being missed. (As already noted, this is equally true of interviews.) Moreover, there is the possibility of an 'observer effect' (Weir and Roberts, *op.cit.*: 173) - also similar to that which can affect interviews - in that those observed may either avoid doing anything that might have a negative effect on the way they are viewed or deliberately set out to please the observer; this effect, Weir and Roberts (*op.cit.*) note, may disappear over a number of observations if the observer is perceived as neutral and is unobtrusive. In observational studies of post-course effects, where there are likely to be constraints on the number and distribution of observations, there is also a risk of sample bias.

3.6.5 Conclusion

Nisbet and Watt (1984: 84) contend that 'interviews reveal how people perceive what happens, not what actually happens'. On the face of it, observation gives a clearer picture of what happens, especially if recordings are available. We should nevertheless be aware that what is observed may also not represent reality in the sense of normality (see the notes under reliability and validity in Table 3.5) and that a single observation, whatever its purpose, will be particularly suspect. As is the case with other data collection instruments, any conclusions will be more soundly based if they are supported by evidence obtained by other means.

3.7 DOCUMENT ANALYSIS

> Programmes produce their own records.
>
> (Weir and Roberts 1994: 136)

The examination of course-related documents can add a further dimension to programme evaluation. Although this point is acknowledged by a number of writers, especially in

relation to projects (e.g. Alderson 1992), it has received relatively little attention in the L2 education literature compared to other programme evaluation procedures.

3.7.1 Categories of document

Weir and Roberts (*op. cit.*) list the following document-types:

- programme descriptions (aims and objectives, criterial features);
- minutes of meetings, circulars, newsletters;
- teaching records, lesson plans, self-report sheets and teacher's guides;
- teaching materials;
- student work.

Henderson's (1978) list includes:

- planning documents (which may be a broader category than Weir and Roberts' 'programme descriptions');
- letters/application forms from would-be participants and records of their previous in-service activities;
- course timetables;
- course hand-outs;
- memos circulated among staff;
- letters from participants to staff.

To these we might add such generally available data as:

- records of attendance;
- records of progress (e.g. test results; rate of progress through scheme of work);
- evaluation data (including summative reports) on previous courses;

and records, perhaps in a diary form, that are less generally available: 'records of decisions made and the rationale for these, and more subjective data, such as accounts of unexpected difficulties and how these were tackled' (McGrath 1995: 15).

The diaries referred to here are analogous to the kind of log kept by a ship's captain, but, as will be clear from 3.4, the burgeoning diary literature encompasses diaries kept for a range of purposes by language learners and teachers in various roles. These can 'help to document ... the everyday working experience that might otherwise be lost' (McDonough 1994: 64).

3.7.2 Purposes of document analysis

All of the more familiar documents listed above can help to fill out the picture in a range of evaluation contexts. For instance, the examination of application forms may make it clear that a programme has attracted applicants of a particular type. This information could inform evaluation in a number of ways. If the intended intake were broader or different it might prompt scrutiny of promotional literature and marketing strategies. Or participant expectations, as expressed in the application forms, could be compared with actual programme content and processes to see what attempts had been made to match the latter to the former. Similarly, the analysis of minutes, memos or briefing documents such as teacher's guides may be revealing as far as leadership styles or communication strategies are concerned; this can then inform changes while a programme is in progress or with hindsight help to explain specific difficulties or the absence of these. On award bearing programmes, data relating to participant assessment can also be examined for a variety of purposes (see 3.7.3, below).

Weir and Roberts (*op.cit.*) see document analysis less as a supplementary source of information than as a check on data already gathered: 'This type of quantitative information can either confirm or challenge insiders' perceptions of the programme, in a depersonalized manner' (Weir and Roberts, *op.cit*: 28). The contrary suggestion made here (and in subsequent sections) is that, appropriately exploited, certain kinds of programme document can also be a valuable source of information in their own right. The next section offers an illustration of this point.

3.7.3 Participant assessment

Assessment, in the sense of formally grading participants or their work, is normally a feature only of award bearing courses. Its most obvious function is to determine the level

of achievement of individual participants relative to the criteria for success on the course. Although there is no evidence in the literature on INSET to suggest that this actually happens, the grades awarded on award bearing courses (and other information relating to participant assessment) can also be used for the purposes of programme evaluation, in one form of document analysis. Let us take five examples.

1. The marks of a whole cohort on a particular assignment are lower than anticipated and closer inspection reveals that participants have been inadequately prepared (the implications of the assignment have not been fully grasped or the bibliographic references are limited); subsequent probing concerning the latter points to problems with library borrowing arrangements.

2. The performance over the course of one or two individuals indicates that they lack the linguistic or academic capacities to cope with the demands of the course. This prompts an analysis of the performance of all participants, with a view to considering the extent to which that performance, successful or otherwise, might have been predicted, or indeed whether information with more predictive power could have been obtained at the admissions stage.

3. Analysis of the grades awarded by different tutors suggests that they are interpreting criteria in slightly different ways and demonstrates the need for a tutor meeting and/or more highly specified criteria. (Problems with tutor grading can also be identified by means of what Parker and Graham call 'Student response slips' (1995: Appendix 1), on which participants indicate whether they understand a tutor's comments on their work and agree with these and the grade awarded - see Appendix 3.6.)

4. Analysis of exam questions (or assignment topics) attempted by participants suggests that certain topics are being avoided by the majority. This may indicate that these topics have been inadequately covered in the course, that one or more tutor-supervisors are less helpful than others, or that tutors' marking standards differ. Further investigation would obviously be necessary.

5. Comparison of participant performance over a period of three or more years reveals that there appears to have been a general decline in standards. This might necessitate a

fairly wide-ranging review of participant profiles, admission procedures, changes in course content or assessment procedures, and so on.

Each of these points might of course be picked up by an External Examiner and, as was acknowledged in the Introduction to this chapter, discussions with the External and the latter's report make an important contribution to the evaluation and development of award bearing programmes. The larger issue is whether programme organisers have established internal mechanisms for scrutinising the data relating to participant assessment and reviewing existing mechanisms.

It is perhaps worth adding that certain types of assignment may also encourage participants to make evaluative comments on a course or be used for that purpose. Participant journals are an obvious example, and Broady (1995) refers to a rather similar concept, the Personal Portfolio. This was one element in the assessment of candidates for the RSA Diploma in the Teaching of Foreign Languages to Adults (by distance learning) under a centre-specific scheme devised by a consortium of institutions led by the Institute for Applied Language Studies, University of Edinburgh (1988-93). The Assessment Guide (IALS, Edinburgh 1991: 23) states that the Portfolio should include a minimum of five written items and tells candidates that it is 'an opportunity for you to ... illustrate your own development as a language teacher [over the period of the course] and the experiences that have contributed to your development' and Broady confirms that it 'encouraged participants to reflect throughout the course on their own learning, on its ups and downs and on its breakthroughs' (1995: 59). However, as is evident from the quotations scattered through Broady's paper, the Portfolio also produced comments that were of direct value in identifying participant concerns and shortcomings in the programme.

3.7.4 Advantages and disadvantages of document analysis

The major advantages and disadvantages of document analysis are shown below, in Table 3.6:

Table 3.6: Features of document analysis

	ADVANTAGES	DISADVANTAGES
economy	low cost	
resource requirements	no special resources required	
ease of administration		dependent on degree of

		centralisation of record-keeping and cooperation of programme personnel
reliability		records may be incomplete or inaccurate
validity	analysis starts from explicit statement of what course is intended to be or has been	records may be carefully edited or falsified; certain types of document (e.g. course report) subject to bias
comparability and generalisability	if standard records available in quantified form, comparison of courses and programmes possible; generalisable statements about tendencies also possible	
utility	evidence of what should have happened and has happened, possibly with explanations	

There would, of course, be a number of problems if we were to rely too heavily on programme documentation. Certain categories of document may be incomplete or missing altogether. The 'records' may also be unreliable or invalid - inadvertently wrong, edited to protect or faked to please (McGrath 1995: 16). Where record-keeping has been imposed on personnel and is either intended to be a measure of control or is perceived in this way, defence mechanisms are likely to come into operation; even where it is not, the enterprise may be undermined by resentment of the additional work involved. Inferences therefore normally need to be treated as hypotheses and cross-checked or corroborated through other means, such as interview.

3.7.5 Conclusion

Document analysis is convenient and feasible in the context of externally evaluated projects, but poses certain problems for institutions committed to end-on courses, where there will be pressure as soon as one course has ended to think about the next, and any decisions regarding further development will tend to be based on the most salient aspects of participant feedback or tutor dissatisfaction. In other words, time will seldom be available, in such circumstances, for the kinds of follow-up and close scrutiny implied in section 3.7.3, at least for members of the course team. It is true that in the case of award bearing programmes an External Examiner's report may highlight discrepancies between documents (or between documents and the programme as experienced by participants) that require attention, but the onus of responsibility for programme evaluation - as distinct from the assessment of course participants - still rests with the institution. It is therefore essential

that programmes, and the procedures by which they are evaluated, be the subject of periodic review, irrespective of whether this is an external requirement, and one element in the review of award bearing programmes it is suggested, should be the kinds of analysis suggested in 3.7.3.

3.8 CONCLUSION

Chapters 2 and 3 have offered a preliminary answer, based on the literature of several related fields, to the first of the research questions stated in Chapter 1:

> *What methods and techniques are available for the evaluation of INSET programmes for language teachers ?*

Chapter 2 discussed the relationship between evaluation foci and method (2.4) and between evaluation stage and method (2.6). In the present chapter, method has been the starting point and the main concern has been to compare methods using a standard set of criteria. At this point, it is relevant to refer to the table of L2 INSET evaluation studies in Appendix 1.5, reworked as Table 3.7, below. Limited though it is, the evidence from the studies surveyed suggests that there is little divergence from the received wisdom of questionnaire, interview/group discussion and observation as the basic methods of enquiry and that the dominant methods of data collection are questionnaire and individual or group discussion. Column 2 shows the number of times the method was used in the 13 studies surveyed.

Table 3.7: Method use in L2 INSET evaluation studies

METHOD	USE	STUDIES
questionnaire	8	Alderson 1985; Cumaranatunge 1989; Palmer, C. 1992; Palmer, G. 1992; Morrow and Schocker 1993; Richards 1995; Ward *et al.* 1995; Wallace 1997
interview	5	Murphy O'Dwyer 1985; Morrow and Schocker 1993; Lamb 1995; Richards 1995; Ward *et al.* 1995
group discussion	4	Palmer, G. 1992; Morrow and Schocker 1993; Richards 1995; Wallace 1997
observation	3	Weir and Roberts 1994; Lamb 1995; Ward *et al.* 1995
document analysis	2	Richards 1995; Ward *et al.* 1995
other:		
- participant journal	3	Murphy-O'Dwyer 1985; Palmer, C. 1992; Palmer, G. 1992
- participant plan	2	Alderson 1985; McGrath 1986
- tutor meetings	3	Palmer, G. 1992; Morrow and Schocker 1993; Wallace 1997
- posters	2	Morrow and Schocker 1993; Wallace 1997
- simulation	1	McGrath and Altay 1990
- written retrospective accounts	1	Murphy O'Dwyer 1985
- informal monitoring	1	Richards 1995
- incidental post-course contact	1	Alderson 1985
- critical discussion	1	Cumaranatunge 1989
- pupil language tests	1	Weir and Roberts 1994

In Chapter 4 we consider how far these public accounts of evaluation practices match the methods generally employed by UK institutions for the purpose of INSET programme evaluation.

CHAPTER 4

SURVEY

4.1 INTRODUCTION

To determine the nature of the evaluation processes current within UK institutions, a survey was necessary. This took two forms:

1. a **questionnaire-based survey** of a sample of institutions offering in-service (mainly TEFL) programmes;
2. **in-depth interviews** with individuals from 'key' institutions, such institutions being those known to offer a number of different INSET programmes and therefore manifesting a commitment to teacher education.

4.2 OBJECTIVES

The principal aims of the questionnaire were to establish **which procedures were typically used** to evaluate in-service language teacher education provision within UK institutions and **the extent to which these procedures were judged adequate** by authoritative informants within the institutions concerned. The results would, it was hoped, offer some insight into the current state of evaluation activity in relation to programmes of this kind, in particular the distribution of evaluation procedures across programme types and institution types (a quantitative question), and the level of awareness among providers of what is desirable and possible (a partly qualitative question). The interviews were intended to flesh out the data obtained via the questionnaires, specifically in relation to informants' attitudes to the evaluation of programmes for language teachers and any unusual features of the evaluation procedures used within their institution.

4.3 SELECTION OF SAMPLE

A variety of considerations influenced the decision as to which institutions should be asked to supply data on their in-service teacher education activities.

1. To take account of the possibility that attitudes and approaches to programme evaluation might differ between public and private sectors and across institution-types, a master list of institutions offering INSET programmes for language teachers was drawn up using The BATQI Register 1993, the BASCELT Handbook (1992-93) and the ARELS/FELCO list. From this, 68 institutions - universities, colleges (of further and higher education and church colleges) and private language schools - were selected following the procedure indicated in 2-5 below. It was hoped that these would yield descriptions of at least 100 programmes.
2. My own experience in the field enabled me to select from the master list a small number of institutions known to be especially active in teacher education.
3. To these were added, again from the master list, a number of institutions in which I had personal contacts (making the possibility of a return more likely).
4. Although a basic criterion for inclusion was the number of programme types offered (the more the better), certain institutions were included simply because they offered a programme of a certain type (e.g. Cambridge/RSA Dip. TEFLA), even if that were the only form of INSET offered.These institutions were selected at random from those offering this type of programme.
5. The relative level of teacher activity within each institutional sector - as revealed by the lists referred to in (1) - was used as a rough yardstick by which to determine the approximate pattern of questionnaire distribution across the three types of institution. (See Table 4.1 in 4.4.5, below.)

The questionnaire-based survey was carried out in 1993. Subsequently, five institutions were selected for special study. With one exception, these were institutions offering a range of programme types, the exception being an institution which stood out after the analysis of the questionnaire data because of its involvement in post-course evaluation. In each case the person with specific responsibility for teacher education was asked if he or she would be willing to be interviewed about the institution's evaluation procedures in respect of in-service programmes. All the potential interviewees were known to me

personally or by reputation. The interviews were carried out between late 1993 and early 1995.

Preliminary analysis of the questionnaires revealed only one institution offering a course for teachers of languages other than English; and this course, the RSA Diploma in the Teaching of Foreign Languages to Adults, as its title implies, is aimed at teachers in further, higher and adult education. Teachers in schools appear to be catered for primarily by Local Education Authorities in England and Wales and their equivalent in Scotland. Although a full survey of LEAs was by virtue of its scale beyond the scope of the research, there seemed some value in including an informant from this sector in the interview sample and since I was based in Edinburgh, the informant selected was the then Languages Adviser for the Lothian Region. Most of the data deriving from this interview has no direct bearing on the survey of institutions reported in this chapter and it is therefore included as an Appendix to Chapter 7, to which it has particular relevance.

4.4 THE QUESTIONNAIRE

4.4.1 Introduction

Data was collected by means of two instruments, the Courses Checklist (Appendix 4.2), on which respondents were asked to list and categorise the in-service courses they offered, and a questionnaire (Appendix 4.3) relating to the evaluation procedures used on these courses. These were accompanied by an explanatory letter (Appendix 4.1). The rationale for each of these documents is discussed below.

4.4.2 The letter

Considerable thought went into the content, style and appearance of the letter accompanying the questionnaire since it was assumed that this might be a critical factor in influencing the recipient to respond to the request for information. The letter:

- states the focus and extent (EFL and Modern Languages) of the research;
- offers the reassurance that permission will be sought if it seems desirable at the reporting stage to refer specifically to a responding institution;
- describes the enclosures and indicates how long it will take to complete the questionnaire;

- specifies a deadline;
- includes an opt-out clause.

Bold type was used to highlight key points (the focus of the survey, the deadline). No attempt was made at concealment; in fact, it was felt that to reproduce the letter on headed institutional notepaper and to include my institutional title might contribute to the intended overall impression: that this was a serious academic enquiry.

4.4.3 The Courses Checklist

The Courses Checklist was designed to gather basic information economically. This information fell under two heads: *identifying* (name of institution, name of informant, telephone number) and *describing* (course type). It was intended that it should:

- provide a relatively undemanding and uncontroversial lead-in to the more specifically focussed questionnaire;
- facilitate classification of courses (e.g. closed-group, open-entry) and enable comparisons to be made of the procedures used on different course types;
- as indicated in the letter, permit those respondents with little time to opt out of completing the questionnaire (in favour of a telephone interview).

Care was taken to reiterate at the top of the checklist the focus of the study ('courses ... for **practising** teachers of EFL or other languages') and to duplicate it in colour so that it was clearly distinguishable from the multiple questionnaire sheets, which were on white. To guard against the letter becoming detached from the checklist, my name and contact address were included. The same precaution was taken with the questionnaire.

4.4.4 Design of the Questionnaire

Respondents were asked to complete a separate questionnaire, a two-sided sheet, for each course.

The questionnaire falls into six main sections, four of which (2-5) relate to the procedures used in different temporal phases of a course (pre-course, in-course, end-of-course and post-course). In each of these four sections, respondents were asked to tick which of a number of different procedures were used. The procedures listed derive from a consideration of the literature supplemented by my own experience. Respondents were

encouraged to add to the list as appropriate; these additions are discussed in section 4.7.2, *Evaluation of the questionnaire*.

What sections 2-5 have in common is that all depend on initiatives taken within the institution concerned; that is, they imply conscious decisions and action. Section 1, by comparison, refers to the use made on award bearing courses of data which is available in the normal course of events, i.e. external performance indicators such as examination results and the reports of external examiners. The final section uses a YES/NO format to ascertain whether the respondents are satisfied with the evaluation procedures used in their institutions and, if not, the reasons for dissatisfaction.

The questionnaire is based on a number of well-attested principles:

- sequence items in a logical order;
- keep rubrics simple;
- gloss any terms that may not be understood;
- keep completion-time to a minimum by requiring only a minimal response;
- be consistent in the use of question/answer formats;
- include open slots in multiple-choice questions to allow for unexpected responses;
- use a clear, uncluttered layout;
- facilitate data analysis by layout, numbering, etc.;
- keep the questionnaire as short as possible.

For a *post hoc* evaluation of the questionnaire, see section 4.7.2.

4.4.5 Return rate

Questionnaires were distributed to a total of 68 institutions and responses, five of which were null, received from 35 (51%). Although this is not a high return rate, it compares reasonably well with those reported elsewhere (e.g. Henderson 1978 refers to response rates of from 10% to 45% for educational surveys). Interviews were later secured with key informants from one of the institutions failing to make returns, who subsequently completed a questionnaire, and from one of the institutions making null returns. This

brought the total number of returns up to 36 (53% of the original list of 68) and completed returns up to 31.

Table 4.1 shows the breakdown of distribution and returns across the three institution types. The null returns are discussed in 4.4.6.

Table 4.1: Breakdown of questionnaire returns by institution type

Type of institution	Questionnaires distributed	Completed returns	Null returns	Ratio of returns
university	26	16	1	17/26
college	23	5	1	6/23
language school	19	10	3	13/19
TOTAL	**68**	**31**	**5**	**36/68**

4.4.6 Non-returns and null returns

Various reasons can be advanced for the non-return of almost half of the questionnaires, the most likely being that recipients - however well intentioned - were simply too busy to spare the time. One of the interviewees pointed out (informal communication) that the questionnaire may not have reached the person best equipped to deal with it because they were away (as was his case) and/or because it was not passed on.

The null returns are interesting because they indicate that there were other reasons why institutions failed to make returns.

Three respondents (one each from a university, a state college and a private language school) said that they had no in-service courses for teachers. Respondent Q, from a university, replied:

> We don't do in-service work in this area. We do do a PG Dip.TEFL and a linked M.A. by dissertation. The latter is done by teachers who have some experience and is evaluated by the usual mixture of internal and external examiner.

One assumes that in this case the PG Dip TEFL is a pre-service programme. Such responses suggest that the mere fact that institutions are featured in one of the lists referred to in 4.3 above as offering courses for teachers is no guarantee that they still do so, that advertised courses actually run, or that practising teachers are the target audience. This

may be a partial explanation for the fairly large number of non-returns from the state colleges and, to a lesser extent, the private language schools.

The Director of one of the most active private language schools added a note to his null return saying that he was unwilling to divulge information which might be of value to competitors (but offering to help in any other way); a similar point was made in a spoken communication by a key person in one of the best-known state colleges, although a further approach led to an interview being granted in the latter case.

4.4.7 Sample Bias

Of the 31 institutions eventually supplying completed returns, four provided data on a total of 52 different programmes (47% of the overall total), and seven each provided data on only one programme. It should be remembered, however, that 'one programme' can cover a wide spectrum of programme-types from one-week open-entry to Master's degree.

There are at least two ways of looking at these figures. One can, for instance, argue that the preponderance of data from such a small subset of the sample inevitably biases the sample as a whole. Or one can say that this very range - from institutions for whom in quantitative terms teacher education is a major concern to those for whom it bulks less large - is in itself representative of the range from within which the sample was chosen. And if it appears that certain forms of evaluation are not being used to any degree across this range of institutions, one can infer that it is highly unlikely that they are used with any frequency in other institutions.

4.4.8 Reliability

In a small number of cases, where two or more named programmes were apparently evaluated in exactly the same way, respondents economised by using a single return for these programmes. Although there is some doubt as to the absolute accuracy of such returns as far as the individual programmes are concerned, any differences were presumably not considered significant by the respondents, and the same information was therefore registered for all the programmes covered by the return.

4.5 THE INTERVIEWS

4.5.1 Planning for the interviews

Planning for the interviews involved consideration of the following:

- whether they needed to be face to face (potential interviewees were scattered all over the country);
- their content and structure;
- how the resulting data would be recorded (e.g. field notes, audiorecording);
- how the potential informants should be approached.

Despite the logistical difficulties, the arguments for face-to-face interviews seemed to be overwhelming and in the interests of speed and efficiency potential informants were contacted by telephone. The purpose of the research was explained to them if they had not already completed a questionnaire return, and they were asked for a face-to-face interview. All those asked for an interview consented.

4.5.2 Conducting the interviews

At the outset of the interview, which often took place several weeks after the telephone contact, the focus of the research was restated. Informants were then asked if they minded being recorded and told that following the interview they would be sent a rough transcript which they could correct or censor. Since there were no objections, all the interviews were recorded. A Sony TCS 2000, a small, good quality cassette recorder, was used in battery mode to obviate any difficulties with leads or sockets. Although conditions were not always ideal (one interview took place over dinner in a noisy restaurant; another was fitted into the gap between trains arriving and departing), informants cooperated fully and satisfactory recordings were achieved.

The interviews went as planned, taking an average of one hour. Questioning started on a fairly general level, either by summarising the questionnaire return or eliciting general information about the types of programmes offered. Given the range of institutions and the evaluation procedures used (key factors in the selection of institutions for more detailed study), no attempt was made to standardise questions and these varied according to the

nature of the institution represented by the informant, the programmes offered, or specific points arising from an analysis of the return or the general findings. The ultimate objective was to encourage informants to talk freely about evaluation policy and procedures within their institution and their own attitudes to these.

4.5.3 Follow-up to the interviews

Following each interview, a transcript was made of those sections thought to be most relevant (5-10 pages of single-spaced A4) and this was sent to the informant for correction and comment. It was made clear that informants were at liberty to censor any sensitive information and that they and the institution could remain anonymous if they wished. A small number of changes and comments were made and since the majority of informants wished to remain anonymous, the decision was made to remove all items which might indicate the source of the quotations used in subsequent sections. The edited transcripts are included as Appendix 4.6.

4.5.4 Analysing the data

The modified transcripts were then subjected to careful scrutiny in respect of the following:

- corroboration of questionnaire data;
- explanations and amplifications in relation to specific features of that institution's questionnaire return;
- the informant's attitudes to specific issues;
- new perspectives.

Relevant quotations appear in the next section. For a critical evaluation of the interviews, see section 4.7.3.

4.6 FINDINGS AND DISCUSSION

4.6.1 Introduction

Following a brief note on the classification of the data, this predictably rather lengthy section falls into two distinct parts. The first deals with the **evaluation procedures used** by

the institutions surveyed (4.6.3 - 4.6.6), the second with respondents' expressed **attitudes to evaluation policy** within their own institutions and to specific evaluative procedures (4.6.7). Both parts draw on questionnaire and interview data. Questionnaire respondents are referred to as A (etc.) and interviewees as 1 (etc.). A reference to e.g. 1/T indicates that informant 1 is also respondent T. There is no significance in the ordering, chronological or alphabetical, of sources.

4.6.2 Classifying the data

It will be recalled that the data gathered by means of the questionnaires, and supplemented by the interviews, comes from institutions of three basic types (universities, colleges, and language schools). For the purpose of analysis, the programmes offered could be classified in a number of ways: award bearing/non-award bearing and closed-group/open-entry being possible broad divisions. The classification actually adopted makes use of these distinctions but also makes a set of finer distinctions within the category of award bearing programmes. Three of these (Master's, B.Ed. and Cambridge/RSA Dip. TEFLA) are self-explanatory. The fourth, 'other certificated', covers any programmes other than these which are validated by universities or examination boards. It therefore includes the Cambridge CEELT, the Trinity College Diploma qualification and Advanced Certificates offered by universities. The final category, non-certificated programmes, is further subdivided into closed-group (contract) programmes and open-entry (individual enrolment) programmes.

The basic quantitative data on programme-types can be seen in Table 4.2, below:

Table 4.2: Completed returns, broken down by course type

Programmes surveyed	
• Master's	21
• B.Ed.	5
• Cambridge/RSA Dip TEFLA	10
• other certificated	18
• non-certificated: closed-group	25
• non-certificated: open-entry	31
TOTAL	110

As noted above, in some instances respondents gave what might be termed a 'generic' return. That is, they used a single proforma for (a) several programme-types, each of

which is evaluated in the same way or for (b) several programmes of the same type (e.g. unspecified advanced certificates or closed-group courses). The numbers given take (a) into account, since in each case the programmes are named, but cannot take (b) into account, for obvious reasons. Some blurring may result.

As hoped, the data covers more than 100 different programmes. These break down into four groups roughly comparable in size:

1. degree-level: 26 (23.6%)
2. other award bearing: 28 (25.5%)
3. non-certificated closed-group: 25 (22.7%)
4. non-certificated open-entry: 31 (28.2%)

4.6.3 Evaluation stages and procedures

A full tabulation of the procedures used by the institutions covered in the survey can be found in Appendix 4.4. This includes procedures other than those specified in the questionnaire (i.e. respondents' additions). In this and subsequent sections specific aspects of this data are examined such as the procedures most commonly used and least commonly used and the degree of variation across programme-types. The findings are summarised in 4.6.3.5 and 4.6.4.5 and possible explanations offered in 4.6.6.

We begin by looking at the global picture in relation to the different phases in the evaluation process, as reflected in the structure of the questionnaire. Numbers in brackets below relate to particular items in the questionnaire in Appendix 4.3.

4.6.3.1 Collection of baseline data

The justification for the inclusion in the questionnaire of certain of the items under this category is that, as indicated in Chapter 3, they can be used in conjunction with other data (e.g. results, measures of participant satisfaction) to assess the adequacy of admission procedures (items 1.1 - 1.4) or the information made available by sponsors (1.5). They can also provide baseline data for a comparison of the effects of the course (1.2, 1.4).

On the evidence of the questionnaires and interviews, before and after comparisons of this type are not normally made. Observation of participants is carried out prior to the

commencement of a course in only 7 out of 110 cases (6%) (Appendix 4.4), and although information about participants is commonly gathered, typically by means of questionnaire (in 48% of cases), interview (35%) and tests (35%), less emphasis is given to this individualised data than to tutors' experience of previous groups (77%), information from sponsors (47%) and tutors' experience of participants' home countries (41%) (Appendix 4.4). What this suggests is that pre-course data collection primarily serves needs other than evaluation. For instance, there may be a bureaucratic requirement or the information is intended for selection purposes or feeds into needs analysis related to course design. Indeed, one respondent explicitly notes against some of the items in this section of the questionnaire 'not done formally as evaluation but as admission procedure' (original emphasis). And this interpretation is supported by interviewees. In the extract below and all subsequent quotations the interviewer's words are in italics and the number in brackets refers to the utterance number in the edited transcript in Appendix 4.6.

> *What about pre-course evaluation, needs analysis and that kind of thing?*
>
> Erm there's a big difference between open-enrolment courses and closed courses. The open-enrolment courses we get papers on all the participants, the application papers the degree certificates the c.v.s and so on and erm they provide a starting-point but we find very often we're more confused than we would be without them because everybody comes from such a different background erm so we use we may very well use that information from before the course to open up the debate about different needs on the course. The closed group type of course it's extremely valuable to have pre-course evaluation available to help us plan ....
>
> (Informant 2: 31)

For Informant 2, then, pre-course information is seen as at best a potentially useful starting point for pre-course planning or negotiation of course content rather than as baseline data for programme evaluation. Informant 3 is even more dismissive of such information, preferring to trust his own judgement of individual development rather than carrying out any more formal comparisons based on pre-course documentation.

> *When people come to you they'll normally fill in an application form [Hm] or they'll write a letter or both of these things [Hm] ... Do you ever go back to that information ... when the results are published ? Do you use it in that way ?*

> Erm not to any significant extent, I would say.... I think that I really find it quite hard to to to make much out of those forms initially ... I find that I learn about the participants from the participants and I don't really feel the need I keep comparing that picture that is unfolding [Hm] erm from them directly erm with previous ideas and future possibilities rather than going back to the paper which I find you know is often not very helpful in in even forming initial impressions [Hm]. Erm I think the only exception to that would be - this hasn't happened for a long time - erm if there was a clear discrepancy between an English language test score on that sheet and a student's performance I think I'd go back to check on that sort of thing [Yeah, yeah] but otherwise very rarely [Right] we do that. That's not to say we shouldn't do it but you know maybe we should be doing so but in practice it's it's very rare.
>
> (Informant 3: 6)

These quotations support the impression gained from the questionnaires that at the outset of a course the main emphasis is on planning for the group (hence the importance attached by survey respondents to experience of similar groups and tutors' experience of the country concerned) and that in planning for a group the information that is available on individuals can be a distraction. Of particular relevance to the present discussion is Informant 3's comment that in his institution they rarely refer back to pre-course information on individuals for purposes of comparison and evaluation, relying instead on evolving impressions. While this is probably not unusual, impressionistic judgements have the disadvantage that they are not recorded and without any record of an individual's starting point there can be no basis for objective comparison at later stages of the programme or for developmental comparisons across individuals.

4.6.3.2 In-course monitoring

Table 4.3 shows the comparative figures (based on the options available in the questionnaire) for formal in-course monitoring for each of the four groupings of programme type (Master's/B.Ed; RSA/Cambridge Diploma/other certificated (Dip/OC); non-certificated closed group (NC cl); and non-certificated open-entry (NC op)):

Table 4.3: In-course evaluation procedures used, showing rank order by frequency

	M/B.Ed	**Dip/OC**	**NC cl**	**NC op**	**TOTAL**
	n = 26	n = 28	n = 25	n= 31	**n = 110**
formal in-course monitoring					
• discussion with whole group	23 (2)	21 (3=)	20 (2)	20(1)	**84 (76%)**
• participant questionnaire	19 (4=)	18 (5)	19 (3)	17 (2=)	**73 (66%)**
• individual participant interviews	21 (3)	22 (2)	12 (4=)	17 (2=)	**70 (64%)**

• tests, exams, written assignments	24 (1)	27 (1)	9 (6)	4 (6)	**64 (58%)**
• observation of sessions	6 (7)	13 (6)	12 (4=)	16 (5)	**47 (43%)**
• meetings of course tutors	19 (4=)	21 (3=)	22 (1)	17 (2=)	**79 (72%)**
• regular written tutor reports	15 (6)	9 (7)	7 (7)	2 (7)	**33 (30%)**

Although there is some variation in the ranking of procedures across programme types, such procedures as discussions with the whole participant group (used on 76% of the courses surveyed), meetings of tutors (72%), individual participant interviews (64%) and - for award bearing programmes - various forms of assessment of participants (58%), all figure prominently as formal monitoring instruments. Observation of sessions is used on approximately half of all programme-types other than Master's/B.Ed, where it is noticeably unusual (mentioned in only six out of 26 cases). On the other hand, written tutor reports, which may include feedback on assessed work, are used more often on Master's programmes (15/26) than on other types of programme (18/84). A variety of other measures are also reported by respondents. These can be classified as:

- other forms of staff-student contact (e.g. student representatives, staff-student committee);
- other ways of evaluating participants;
- the involvement of people outside the course team.

Although the questionnaire did not refer to informal evaluation, several of the interviewees stress its importance. In-course informal feedback breaks down into the comments volunteered by course participants (during coffee breaks and tutorials) and incidental evidence of the sort referred to in the quotation below, which can be picked up almost unconsciously by an experienced tutor.

> I mean the (laughs) the very first session especially on the French course my colleague erm who deals with that course and has done I think since it started about twelve years ago is able quite unerringly in the first session (laughs) to spot (laughs) certain tendencies among certain participants which which sure enough you know develop er just as she predicts erm during the ensuing three weeks.
>
> (Informant 3: 4)

Such data is likely to be available in greater quantity over longer courses, especially if the atmosphere within the institution is conducive to frequent informal contact between course participants and staff. However, judging by the testimony of Informants 2 and 3 (see the quotations below), it appears that there is a greater tendency to attend to informal data

when courses are short. Constraints on the amount of evaluation that can be done on short courses can even mean that informal evaluation is tacitly accepted as a substitute for more formal measures.

> *Can we go to ... the the two-week courses that you run ? [Yeah.] Do you evaluate these courses at all ?*
>
> Yes, Just about. An end-of-course questionnaire. Erm the other way that we do it is that I suppose there's a lot of informal evaluation going on all the time with the participants on the course in that we have a pretty open-house policy at [INSTITUTION] where participants are encouraged to keep talking to the tutors about what's going on on the course ...
>
> (Informant 2: 20)

In the case of short repeat closed-group courses, evaluation may even be tokenistic:

> ... summer courses ... tend to be closed-group ones, two to three weeks in duration and with them they all have an end-of-course evaluation but if if they were only two-week ones like the [COURSE] I did this year we didn't do any evaluation formal evaluation except right at the end but the group we have from [COUNTRY] that I've also been involved with that's a three-week course and we've usually done a brief evaluation at the end of the first week and then at the end of the course a more thorough one. But just given the shortness of them and also the fact that erm since they are courses which are which have been going for some time with a very similar clientele [*Yeah*] there's not the feeling of I think of needing to dig you know much deeper than that er and so on to in terms of our evaluation procedures...
>
> (Informant 3: 3)

Both informants appear to set great store by informal evaluation. The value of informal feedback is that it may be more honest than what is said in open forum, but it may equally be both unrepresentative and unreflective. While a potentially useful complement to data gathered by other means, it is not a substitute for more formal data-collection.

The constraints associated with short courses are, of course, real, and it is reassuring to feel that participant groups are so similar from year to year that few changes need to be made to a programme and that the success of the course can be more or less taken for granted on the basis of previous evaluations. And yet without constant checking, possibly selective and possibly using different methods, there is no certainty that participants from year to year really are so similar in their level of awareness, skill and wants (which is a

further argument for baseline data collection), that the course which they following is delivered in the same way, or that their reactions are similar to those of previous groups. Repetition does not remove the need for evaluation.

4.6.3.3 End-of-course evaluation

Table 4.4 shows the comparative figures (based on the options available in the questionnaire) for end-of-course evaluation for each of the four groupings of programme type.

Table 4.4: End-of-course evaluation procedures used, showing rank order by frequency

	M/B.Ed	Dip/OC	NC cl	NC op	TOTAL
	n = 26	n = 28	n = 25	n = 31	n = 110
end-of-course evaluation					
• discussion with whole group	16 (4)	19 (3)	21 (2=)	22 (2)	**78 (71%)**
• participant questionnaire	23 (1)	22 (1)	23 (1)	25 (1)	**93 (85%)**
• individual participant interviews	15 (5)	7 (6)	7 (5)	7 (5)	**36 (33%)**
• tests, exams, written assignments	21 (2)	14 (4)	5 (6)	- (6)	**40 (36%)**
• meeting of tutors	19 (3)	21 (2)	21 (2=)	17 (3)	**78 (71%)**
• written tutor reports	7 (6)	11 (5)	18 (4)	6 (4)	**42 (38%)**

On this evidence, the most commonly used end-of-course evaluation measures are participant questionnaires, discussion with the whole participant group and meetings of tutors, all of which are used on more than 70% of the programmes surveyed. Written reports figure, especially in the case of closed-group programmes, where they are used on 18 of the 25 programmes; external examiners' reports were also mentioned in the returns in relation to ten of the award bearing programmes.

4.6.3.4 Post-course evaluation

As can be seen from Table 4.5, below, the questionnaires reveal a striking lack of post-course evaluation.

Table 4.5 Summary of post-course evaluation procedures used

	M/B.Ed	Dip/OC	NC cl	NC op	TOTAL
	n = 26	n = 28	n = 25	n = 31	n = 110
post-course evaluation					
• participant questionnaire	2	2	5	3	**12 (11%)**
• report from sponsor or other responsible person	3	1	8	-	**12 (11%)**
• observation of participants' teaching	1	1	4	-	**6 (5 %)**
• follow-up meeting for participants	-	1	7	-	**8 (7 %)**

Of the 110 programmes described, only about one in ten are followed up in any systematic way (participant questionnaire (11%) and sponsor reports (11%) being referred to most frequently). Of the 58 references to post-course evaluation 30 apply to closed-group programmes, several of which form part of ongoing projects. It is tempting to speculate that the initiative in these cases came from the sponsors, but this is denied by Informant 4, who says, referring to a particular programme: 'because that was a kind of ongoing thing that was developing I felt it was important to find out what effect the course was having' (4: 1), and because she found this procedure useful for purposes of programme development, it 'tended to carry over into our pattern of course evaluation and course design' (4: 1).

In one return, reference was made to the use of 'action plans'; however, further investigation indicated that this phrase related to plans formulated by the providing institution in respect of future courses rather than participant action plans of the sort referred to in Chapter 3.[12]

Many informants are only too well aware of this gap in their approach to the evaluation of their courses: 'The biggest gap is post-course follow-up after a period of time, more than 3 months - perhaps even 4-5 years. The MSc is a major investment by the participants - the return may well not be immediate' (Respondent X). For other similar comments, see *Attitudes to evaluation* (4.6.7).

4.6.3.5 Summary

The findings discussed in the previous subsections are summarised below in the form of a series of statements.

- There is no evidence to suggest that information on participants that is available prior to the commencement of a course is subsequently used for purposes of programme evaluation (4.6.3.1).

- In-course monitoring is both formal and informal. The most commonly used formal measures are discussion with the whole participant group and meetings of course

tutors. On short courses there appears to be a greater tendency to rely on informal data (4.6.3.2).

- A variety of measures is used for end-of-course evaluation: participant questionnaires, formal discussion with the participant group, meetings of tutors, written tutor reports, and external examiners' reports; of these, the first three are more widely used (4.6.3.3).

- Little post-course evaluation is carried out (4.6.3.4).

4.6.4 Programme types and evaluation procedures

Thus far we have been concerned with tendencies across institution types and programme types. In this section we compare the findings for each of the four broad categories of programme type proposed in 4.6.2:

1 degree-level programmes: Master's and B.Ed;
2 other award bearing programmes: RSA/Cambridge Dip TEFLA and other certificated programmes;
3 non-certificated closed-group programmes;
4 non-certificated open-entry programmes.

The commentary is again based on Appendix 4.4.

4.6.4.1 Degree-level programmes

Analysis of the procedures used on Master's programmes indicates that most measures are used by most institutions: all but three of the items in sections 1-3 of the questionnaire are used on more than half of the 21 programmes surveyed. Even so, there are a small number of slightly surprising results, such as the fact that individual participant interviews and whole-group meetings are used as a form of end-of-course evaluation on only 13 programmes. One reason for these relatively low numbers may lie in the use by 19 institutions of end-of-course questionnaires. Other reasons may include the size of class on some Master's courses and the fact that a good deal of informal feedback may be received during the course. As one interviewee put it: 'Every tutorial that we hold during a Master's course is an informal evaluation. You always get feedback' (Informant 2: 6).

The B.Ed sample (5) is much smaller and more homogeneous, perhaps because it relates largely or totally to groups from the same country and the courses in this category can therefore be more appropriately thought of as award bearing closed-group courses. In this connection, tutors' experience of the country and of previous similar groups, together with information from the sponsor, all become more important as ways of assessing pre-course knowledge, skills and wishes. In relation to in-course monitoring, apart from regular tutor meetings, there appear to be no marked differences from the Master's courses; the pattern of end-of-course evaluation is also similar. As on the Master's courses, post-course evaluation is almost non-existent: reports from sponsors are mentioned in just two returns and observation of participants' teaching in just one.

There are three outstanding findings within the subset of degree-level programmes, all of which relate to Master's courses. These are:

results (number and level of passes): if we are to judge by the returns, only seven institutions take cognizance of results in evaluating their courses (*cf* the institutions offering a B.Ed, all of which attach importance to this). This apparent lack of concern for results is a little hard to credit. Results inevitably reflect on an institution (the demands of the course; the support given to students; selection procedures) and if there are failures it is to be expected that an institution will go back to the pre-course information on which their decision to admit these students was based. It could be argued that this process of after-and-before comparison ought to be standard for award bearing programmes, since it would benefit both the institution (by identifying the kinds of student who should either be screened out at the selection stage or carefully monitored) and the individual student (for whom, in extreme cases, rejection might be kinder than acceptance). Surprising though this finding is, it is corroborated by limited evidence from the interviews (see the discussion of pre-course evaluation, above).

observation of sessions (e.g. by CD): this appears to happen on only four of the Master's programmes surveyed (and on two out of five of the B.Ed programmes). It is not difficult to find possible reasons for this finding (see, for example, the discussion of institutional culture in 4.7.4, below), but it is nevertheless somewhat surprising in the present quality assurance climate.

post-course evaluation: with the exception of two institutions - one of which uses participant questionnaires while the other relies on reports from sponsors - the questionnaires recorded a total absence of evaluation activity in terms of subsequent effects as far as the Master's programmes are concerned. We shall be returning to this finding in later sections (4.6.5, 4.6.7, 4.8.2).

4.6.4.2 Other award bearing programmes

One striking feature of the findings relating to the evaluation of the ten Cambridge/RSA Dip TEFLA programmes compared to the Master's programmes is that all the institutions surveyed attached importance to participant results. Since these results are sometimes published and since applicants sometimes ask about them, this is hardly a surprising finding, but it does throw the corresponding finding for the Master's programmes into sharp relief. Moreover, there is no evidence to suggest that organisers of Diploma programmes pay any less attention to other forms of evaluation. Applicants are very carefully screened (interviewed in all ten cases, and in some also sent a pre-course questionnaire and/or observed teaching and/or tested). The level and nature of in-course monitoring roughly parallels that for Master's programmes, although there is a stronger tendency for sessions to be observed (in five of the ten cases) as part of the evaluation process. As a form of end-of-course evaluation, individual participant interviews (2/10) are less usual than whole-group discussion (8/10) or participant questionnaires (6/10), perhaps because in general such courses tend to involve a good deal of formal individual contact between participants and tutors. Post-course evaluation is again very much the exception rather than the rule, as this extract from the interview with informant 5/A explains:

> *The Dip TEFLA ...is assessed externally* [Yes] *and validated externally* [Hm hm]. *Erm how interested are you as far as your evaluation of the course is concerned by the results ? Do you to some extent say 'Well, we must be doing all right because the results are good in our terms'? Do you lean on that as confirmation that you're doing a good job ?*
>
> I think for the most part we do...
>
> *Would you say that because you have this kind of external measure of the success of the course that you feel under less pressure to evaluate the course for your own purposes using your own means ?*

> Yes, I think that's fair. There isn't time [*Hm*] to evaluate as we go along [*Hm*]. Erm I know that sounds (laughs) unscientific but it's perfectly true ... the evaluation that we do ... is more to do with ... working out from the s the s participants how they're erm feeling at any particular stage and whether they need more or less ... So we don't evaluate formally as we go along ...
>
> *And do you carry out any kind of post-course follow-up to try and trace people afterwards ?*
>
> No... (1-3)

The second item in the questionnaire asked whether results were used to evaluate the success of courses, the hypothesis being that external assessment and validation may cause institutions to feel that there is less need for programme evaluation as such. As we have seen, and as Appendix 4.4 indicates, practices on this point vary across institution-types. The extract above suggests that in certain institutions at least the combination of participant assessment during such a course and external assessment do exert an influence of the kind hypothesised.

The second category of programme within this rough grouping is 'other certificated programmes'. As befits this rather mixed bag of programmes validated by universities on the one hand and examination boards on the other, the findings for 'other certificated programmes' tend to fall midway between the Master's programmes and the Cambridge/RSA Dip TEFLA. For example, on 11 of the 18 programmes, results are taken into account in evaluating the success of a course, and pre-course questionnaires are sent out in a number of cases. In general, both the level of evaluation and the procedures used are broadly similar to those on Master's, B.Ed and Cambridg/RSA Dip programmes.

4.6.4.3 Non-certificated closed-group programmes

Table 4.6 below shows the average number of measures used (based on the options available in the questionnaire) for in-course monitoring and end-of-course evaluation for each of the categories of programme type. The questionnaire listed seven in-course measures and six end-of-course measures. In this case, degree-level and other award bearing courses have been kept separate to preserve, most obviously, the distinction

between the B.Ed and other programmes. (The relatively high levels of evaluation for B.Ed programmes may, of course, be partly a product of the size of the sample (5).

Table 4.6 Average number of measures used on different programme types

	M	B	Dip	OC	Nc cl	NC op
in-course (7)	4.8	5.2	4.9	4.6	4.0	3.0
end-of-course (6)	3.6	5.0	3.6	3.2	3.8	2.4
TOTAL	**8.4**	**10.2**	**8.5**	**7.8**	**7.8**	**5.4**

The comparison is particularly revealing in relation to the non-certificated courses. The non-assessed nature of such courses is an obvious factor in their lower averages for in-course evaluation (and other factors are discussed in 4.6.6, below) but what is striking about the figure for end-of-course evaluation of non-certificated closed-group courses is that - if we leave aside the B.Ed programmes - it is marginally higher than that for the award bearing programmes and significantly higher than that for non-certificated open-entry programmes. The reason is not hard to find:

> there've been over the last few years a number of occasions where we've run a course and have wanted to do a repeat course because we've known that the finance is there or the project has a certain life and it's very important for us to get it right next time round if we didn't get it right first time round
>
> (Informant 2:17)

At a more specific level this client-centredness manifests itself in a number of predictable ways. Pre-course information is of particular importance, and pre-course questionnaires (used on 21 of the 25 programmes) and information from sponsors (17/25) figure prominently in the findings, as do tutors' experience of previous similar groups (19/25) and, to a lesser degree, of the country concerned (12/25). In several cases, representatives from an institution have carried out information-gathering visits to the country from which participants come. Although participant assessment is less significant (9/25) on most of these courses, one exception being a course on which participants are given an ARELS (oral) test on arrival and just prior to departure, regular meetings of tutors are a common feature (22/25), sessions are observed in about half of the cases, and a certain amount of report writing by tutors is involved, noticeably at the end of the course (18/25). As end-of-course evaluation measures, almost all institutions use both a participant questionnaire (23/25) and a whole-group meeting (21/25). The highest level of post-course evaluation is reported for this category of programme (30 of a total of 58 references). Questionnaires

are used (5/25) and sponsors' reports also provide feedback (8/25), but follow-up meetings and visits are also mentioned (e.g. meeting for participants, observation of participants' teaching). Given the financial implications, the concern shown for evaluation on such courses is logical.

4.6.4.4 Non-certificated open-entry programmes

Explaining the difference in the amount of evaluation done on non-certificated closed-group and open-entry programmes, Informant 4 says: 'the mixed groups every group's always different and er I suppose when we've got monolingual groups they force you to think a bit harder about what you're doing because well they come from the same backgrounds' (20). Although it is clear from Table 4.6, above, that non-certificated open-entry programmes are in general less intensively evaluated than other programme types the evidence from the questionnaires is that most institutions do evaluate even short programmes, the norm for which seems to be 2-3 weeks. On the 31 programmes surveyed pre-course questionnaires are widely used (23/31), presumably as an input to needs analysis, but this may be a routine procedure since there is also an overwhelming tendency to draw on tutors' experience of previous similar groups (24/31). In-course monitoring tends to be through a combination of formal discussion with the whole group (20/31), meetings of course tutors (17/31), and individual interviews (17/31). On 16 of the programmes there is also observation of sessions (e.g. by the CD). Participant diaries are mentioned as a source of evaluative data by one institution. As far as end-of-course evaluation is concerned, the most popular procedures are participant questionnaires (25/31), formal discussion with the whole group (22/31) and meetings of course tutors (17/31). Several institutions also mention reports, written or otherwise, by the CD. Post-course evaluation is for the most part indirect (e.g. participants return to do another course or recommend the course) or incidental (chance meetings with former participants); participant questionnaires are used in two cases.

Several interviewees comment on the problems associated with evaluating courses of this type (e.g. 3: 7; 5/A: 4-5, 47-48). These problems are discussed in 4.6.6, *Factors affecting evaluation*.

4.6.4.5 Summary

The key features of the findings discussed in sections 4.6.4.1 - 4.6.4.4 can be summarised as follows:

- End-of-course questionnaires are the most common form of programme evaluation across programme types (used on 85% of programmes surveyed). Whole-group discussion and tutors' meetings are also widely used for both in-course and end-of-course evaluation. The extent to which other procedures are used appears to be largely related to programme type.

- Course results are not universally used in the evaluation of award bearing programmes. (This is consistent with the finding that such baseline data as is available is not used for evaluation purposes.)

- In-course evaluation tends to be less intensive on the shorter (non-certificated) programmes.

- Non-certificated open-entry programmes are less intensively evaluated than other types of programme.

- In non-certificated closed-group programmes end-of-course evaluation receives more attention (with the exception of the B.Ed. programmes) than in other programme types.

- There is some evidence of post-course evaluation activity in relation to non-certificated closed-group courses; this is strikingly absent from the evaluation of other programme types.

4.6.5 Less commonly used procedures

The questionnaires provide quantitative information on the procedures used, but the interviews provide a wealth of detail about specific procedures, such as diaries, post-course questionnaires, follow-up visits and feedback from people outside the course team. Since the institutions using these particular procedures are apparently in a minority but the

procedures are felt by the informants concerned to be particularly useful, the relevant sections of the interviews are here reported in some detail.

4.6.5.1 Diaries

Although only one of the questionnaire returns referred to participant diaries, these were mentioned in two of the interviews. This section draws on these interviews to establish how diaries are used for evaluation and what advantages they are perceived to have.

As Informant 1/B (16) makes clear, diary use is widespread in his institution and he is clearly aware of the potential for evaluation.

> *What do you feel comes out of the diaries in evaluation terms ?*
>
> You learn people say what they like and what they don't like and what they feel they need people often say very interesting things like in groups where there are problems in the group they write about the problems within the group and their problems with each other - that's not what you want, no. Erm ... some people write to please and they they just say everything's wonderful and what they like. It's not terribly useful but I don't think there's much you can do about that.
>
> (17)

One closed group wrote about each other, about the tutors and about the course, in very bitter terms: 'they felt that ... they had been misled as to the objectives and content of the course deliberately in [COUNTRY]'. In such situations, he comments, 'you can't always respond ... but it's much better to know what it is that they're angry and unhappy about than just to kind of see an angry and unhappy group of people' (Informant 1/B: 18). The central purpose is, however, to obtain feedback on the course: 'The programme is is centrally determined and the one of the functions of the diary is to to keep a check on how people are reacting to that' (22).

Diaries can also reveal specific needs. If there are

> one or two people whose needs can't be met in the group you can erm "massage" them and give them books to read on topics that they want and arrange for them to go and talk to your colleagues who're particularly interested in the topic. No that's not massaging them that is meeting their needs to some extent.

(22)

Informant 4 also testifies to the value of diaries, though these were not mentioned in her questionnaire return:

> I find that diaries are really useful if you can persuade the students of their value and often I find that a two-week course is not long enough for some individuals to to grasp how useful that can be for us.
>
> (17)

She refers to an occasion when she was preparing a tender for a closed group of the kind she had taught two years previously:

> I found that going back though the diaries actually helped [*Hm*] to bring back memories of what had gone on [*Hm*] and in a way were more useful than the post-course evaluations that we did ... because you could see how things progressed through [*Hm hm*] through the course and also if you've got twelve diaries you get different perspectives on what's actually going on. It would be really quite nice to get teachers to to keep diaries of what what they do when they go back and follow up ...
>
> (17)

Not all participants are willing to cooperate.

> ... they just didn't like reflection I think. [*Right*] And I think this whole business of reflective practice is something which is fairly alien still to a lot of teachers and they kind of see it as intrusion some kind of intrusion [*Hm*] ... I've tried different approaches and sometimes feeding in an article on the value of diaries [*Hm*] - this went down like a dead duck (laughs) last time I used it - erm showing samples of different diaries .... Tried doing sort of group group diaries [*Hm*] ... and again that's sometimes quite interesting to to see how people influence each other in terms of how what they put down [*Hm*] and how they perceive it ... But I must say it's been much more on an ad hoc basis rather than a systematic approach.
>
> (17)

To judge from these comments, participant diaries are used primarily as a supplementary monitoring instrument, their particular value from an evaluation perspective being to reveal individual needs or feelings about the course or other individuals. Useful though

such information may be, the impression given by the interviews is that because diaries - like reflection - are felt to be a good thing pedagogically their contribution to programme evaluation (as an alternative to individual interviews, for instance) - and what that implies for the way they are presented and the data processed and used - has not been fully thought-through.

4.6.5.2 Post-course questionnaires

The general neglect of post-course evaluation has been noted in previous sections, as has the dominance of the end-of-course questionnaire as an evaluation instrument. Informant 2 (7, 11, 8) finds post-course measures in the form of questionnaires a much more reliable way of perceiving effects than an end-of-course evaluation (for an explicit contrast, see the emboldened sections):

> We send out a questionnaire to all of our Master's students in March each year erm to the last year's Master's group when they're back in post and we get very valuable information about what the course has meant to them erm once they're back in post. The reason why that why that is is simply because erm **they er often can't get the course in perspective for themselves until they've returned to their posts and started to think about what the course meant for them**.
>
> [...]
>
> we were very clear right from the very start of that course that it would have been premature to evaluate the course on the last day and so we we've always felt that erm our students wouldn't fully understand what the course meant to them until they got back to their own countries and started doing the things that were expected of them when they returned to post and I think that's been borne out quite often. And **the emotional involvement of the student in a course during a course is so high that you get some very distorted data from mid-course and end-of-course evaluation procedures,** which have to do with personal agendas...
>
> [...]
>
> *Do you feel that this is your most important way of getting data on student satisfaction ?*
>
> On Master's courses, yes.

For further comments on the use of post-course questionnaires by this Informant, see Appendix 4.6, Informant 2: 7-14.

Informant 3 observes that although his institution had used post-course questionnaires in the past to evaluate courses of shorter duration than a Master's, they have now given this up.

> I think we kind of gave up on that because it because of pressure of work I think, just inertia, but also because it didn't seem to be yielding - very low return-rate [*Hm*] - it didn't seem to be yielding a great deal of information.
>
> (5)

He adds that although the institution produces an annual newsletter and this contains a request for news of former course participants, this call invariably falls on stony ground.

Such disparate experiences are explicable in terms of programme-length or level of attachment ('loyalty') to the institution. The way in which the questionnaire is 'packaged' may also be a factor. On these points, see sections 9, 10 and 13 in the transcript of the interview with Informant 2.

Informant 4, whose institution also sends out post-course questionnaires, points to the fact that course participants may need to be persuaded that evaluation is not just a routine institutional exercise:

> I suppose we make quite a big thing about the importance of feedback to us [*Hm*] and that getting questionnaires saying how wonderful we are isn't terribly helpful (laughs) to [*Hm*] the course development ... and ... we're not just going through the motions as it were ... part of the key is trying to impress upon them while they're here how useful evaluation is [*Hm*] by demonstrating how we implement what comes out of our evaluations.
>
> (16)

She also feels that on short programmes (3 weeks) timing is crucial, explaining the 'surprisingly good' rate of returns by the fact that 'we played around a little bit with with the timing' (6): 'if it's too too soon after the course you get you might get returns but nobody's (laughs) tried anything out [*Hm*] and if you leave it till later on in the year when they're really busy marking exams then [you?] don't get returns' (6).

The positive experiences of Informants 2 and 4 suggest that if post-course questionnaires are prepared for, appropriately packaged and timed, and in due course carefully processed, they can be cost-effective way of gathering data on post-course effects.

4.6.5.3 Feedback from people outside the course team

One concern in the literature on programme evaluation is that insiders are not the most objective of evaluators. There is scant evidence from the questionnaires that this is a serious matter for the institutions surveyed. The interviews are more revealing in this respect. Informant 1/B explains that in his institution a course tutor will sometimes ask him to come in if a problem has arisen which the tutor has no authority to deal with; the role of the 'outsider' as troubleshooter is also referred to by Informant 2 (24) and Informant 5/A (20).

In the institutions of Informants 1 and 5, however, the insider-outsider also has a more general role to play in respect of evaluation.

> At the end of every course somebody ... will spend an hour with the course participants erm doing what is disparagingly called a happiness sheet, I think .... they know this is going to happen. In that final week they're asked to reflect on certain criteria and ... typically they will in groups discuss the criteria and appoint a spokesperson [*Hm*] so when for example I go into the room they're primed and ready to speak and I see my role as being erm to say things like "How many people agree with that?" And and I write notes as I listen to them [*Hm*] and then I read the notes back to them and they say "No, actually it's not 'very' it's only 'fairly'". [*Hm hm*] That gets typed up and that record is kept. I would put if there have been problems during the course as there sometimes are ... I would add a record of these problems to that written record.
>
> (Informant 1B: 4)

Asked if he feels that it is important for the discussion to be conducted by someone external to the course team, he says: 'Yes... that is the policy but erm erm it's me who imposes it (6).

Sometimes the procedure is different:

> ... one thing I'll do - I'm thinking of 2-week 30-hour open-access courses - I'll go in and I'll say erm I'll state what the aim of what the objective of the course was and what we hope that they will feel that they've achieved and then say absolutely nothing [*Hm*] and then there's quite a long silence [*Hm*] and then somebody will say something ... and I will play it entirely by ear the only thing that I will really do is to ensure that everybody has the opportunity to say something.
>
> (7)

On a longer course this might happen more frequently (7).

In Informant 5/A's institution, the taking of participant feedback by a senior member of staff has also become the norm: 'it's to do with feelings and not hurting feelings' (23). The discussion with the group would normally be preceded by a consultation with the course tutor 'for the sort of grey areas where you just don't know ... how they felt ... but also to know who are the loudmouths ... There has to be about half an hour's discussion beforehand' (24).

Sponsors can also be important sources of information. Informant 2 acknowledges the dual role that sponsors can play in passing on feedback to the course team:

> that's fairly common erm even during the course that er representatives of the sponsors will call in and actually spend time with the students, with the tutors and evaluate what's going on. It's quite important to do that because er well sometimes the students come from a culture er where deference to the teacher is all and they would say things erm behind the teacher's the tutor's back that they wouldn't say directly to the tutor's face [*Hm*] for fear of giving offence [*Hm*]. And we've seen we've on a number of occasions had valuable feedback from sponsors' visits visits during the course. Erm we also then get er valuable feedback from sponsors when they're back in their own country erm which had to do with the degree to which the course has made a difference. But that's that's usually informal.
>
> (26)

While recognising that sponsors can feed back useful information while a course is in progress, Informant 5/A (11) expresses reservations about the information to be gained from sponsors when the latter are of the same nationality as participants: like the latter, the representative of the sponsors may also hold back information to be polite or for some other reason: 'the feeling has been yes they are telling us how things are because that's

how we think things are but if there were a problem ... we can't be sure we'd be getting that' (50).

Other sources of feedback are also mentioned. In the questionnaire returns, Respondent I makes reference to the fact that a member of another department within the faculty is involved in programme evaluation, and Respondent X notes that the end-of-course meeting with the whole group of Master's course students and individual interviews are conducted by the External Examiner without course tutors being present.

Towards the end of his interview, Informant 1/B offers a differentiated view of what he feels to be the most useful evaluative measures:

> For the quality control ... the summative evaluation at the end when another person is there with the group; for making changes to subsequent courses the post-course feedback ... and for making changes during the course probably diaries mediated with discussion which arises between the tutor and the people which arises taking into account what she or he has read in the diaries.
>
> (23)

It is not really important whether one agrees with these particular purpose-method matches, although they seem well informed. What matters is that institutions actively consider the various purposes for which they need information and how this can best be collected. This will in turn raise other issues of the sort discussed in Chapters 2 and 3.

4.6.6 Factors affecting evaluation

In seeking explanations for the findings discussed in the previous sections, we shall consider programme type, programme length, the time available for evaluation, institution type and institutional culture.

4.6.6.1 Programme-type factors

The **choice** of procedures is clearly related to programme-type, i.e. certain data gathering and evaluation procedures form a normal and expected element in the evaluation of some programmes but not others. Examples include:

- *tests/exams/written assignments* during a course (or at the end of a course): these are a normal feature of award bearing programmes and unusual on a non-certificated (usually shorter) programme unless a sponsor requires this.
- *external examiner's report*: only appropriate on award bearing programmes
- *follow-up meeting for participants*: normally only possible in the case of closed groups or overseas-based programmes.

One feature of all these examples is that they relate either to award bearing programmes or to closed-group programmes. It might be expected that if certain procedures are only used on these specific programme types, then programmes which do not fall under either of these heads (i.e. non-certificated open-entry programmes) will be less rigorously evaluated or at least evaluated differently. Although this may be generally true (as illustrated in Table 4.6) it is not necessarily the case for all programmes, as we shall see in section 4.6.7.3.

The number of measures used may, however, be affected by participant accessibility. One respondent, from an institution which offers a sandwich course, notes that when students are following the face-to-face components 'evaluation procedures are good. Between these periods (i.e. when they are studying on their own) there is room for improvement' (Respondent L).

4.6.6.2 Programme length/time available for evaluation

Judging by the number of comments on these issues, programme length and the time required for evaluation are important factors in the **amount** of evaluation that is carried out.

Respondent 5/A, commenting in her questionnaire response on a two-week programme, says: 'We aim to have individual tutorials on all our teachers' courses. With the two-week open enrolment courses this does not always occur due to time factors'.

In the interview, she expanded on this:

> ... to be honest we've never had the time we're always looking forward. I mean this is the nature of any organisation I suppose [*Hm hm*] and again what's its value what's it going to do what's its purpose? In fact we're just finished and it's finished.
>
> (47)

Alderson's (1985) analysis of the reasons for lack of follow-up, referred to in Chapter 2, clearly has some basis in fact.

About a five-week programme, the same informant writes: 'Could have better evaluation for this course. However, participants also need to be trained in giving evaluation. A five-week course is not long enough to do this adequately'. Time constraints are clearly in the minds of several respondents. Respondent C notes of a five-day programme that it is 'too short for evaluation of progress'; and respondent D's gnomic comments - 'This is a four-week course/This is a two-week course' - imply the same, although he goes on to list a variety of procedures used. Respondent E writes: 'This is a practical workshop leading to no qualification', but again lists a number of evaluative measures used. Similarly, Respondent 1/B states in the questionnaire return: 'When a course is as short as 27 hours, we cannot devote too much time to evaluation', but during the interview acknowledges that this may have been misleading because several measures are used.

The five responses just described fall into three categories: C and D seem to feel that it is impossible (or inappropriate ?) to attempt to evaluate a short course (implicitly defined as one month or shorter), whereas E implies that evaluation is inappropriate for a practically-oriented, non-certificated course. A and to a lesser extent B appear to be concerned about the fact that they cannot do more evaluation on courses of this length. A and B refer explicitly to time, which in the context of their remarks probably relates to competing demands on the time of participants. This clearly lies behind the comment of other respondents: 'I would like to have more student evaluation but time/syllabus get in the way' (Respondent F) and 'too much time given to evaluation is perhaps perceived ... as a bit of a waste of time' (Informant 4: 20). In other words, time intended for student learning cannot justifiably be used for evaluation.

The amount of evaluation is also affected by other demands on **tutors' time**: '[Evaluation] could be tighter, more formal - question of time and money' (Respondent G); 'in general the information is not analysed in sufficient detail (e.g. the outcomes in quantifiable terms)' (Respondent H).

A slightly different perspective comes from Informant 3:

> in a short course it's very difficult to have a major change of direction I mean everything's happening so quickly. You know, by the time you've thought about (laughs) making a change it's too late (laughs) the course has finished erm. I mean, that's not to say it might not be very useful for the next time round but I think ... we do operate on the assumption that or I do anyway that in in the shorter courses erm if it's less than if it's two weeks or less [*Hm*] I suppose two weeks is the shortest in fact [*Hm*] that we have in terms of sort of formal and thoroughgoing and major evaluation it's often unfeasible. That's not to say that there isn't room for plenty of other you know more informal and ongoing fine-tuning type of evaluation.
>
> (4)

- a view shared by Informant 5/A (5, 51).

In some institutions the time factor is more a matter of **the time needed to establish a satisfactory system of evaluation:**

> Not yet developed post-course evaluation (Respondent I).
>
> At the moment I feel we do not have the right balance between individual and group feedback, nor have we yet found the right formats, wording, etc for our evaluation instruments (Respondent J).
>
> Procedures are not yet standardised across Faculty (Respondent K).

The lack of defensiveness in these and the comments quoted earlier in this section suggests that institutions are far from complacent about programme evaluation. However, there is little doubt that the widespread use of end-of-course questionnaires, whole group discussions and tutors' meetings owes much to their convenience as evaluation procedures.

4.6.6.3 Type of institution

It is difficult to make comparisons across institution types since to a certain extent programme types are associated with institution types (e.g. Master's programmes with universities and non-certificated open-entry programmes with language schools). What is evident is that there are differences in the amounts and types of evaluation within institution types. In the present sample, these differences are perhaps more marked in the language school sector than among the universities, although this may be partly because, as noted above, award bearing programmes carry with them certain imposed evaluation requirements.

4.6.6.4 Institutional culture

It would be surprising if there did not exist a degree of variety. Institutions vary not only in type but also in 'culture' (i.e. beliefs, and the norms and behaviour patterns associated with them). Within a university, for instance, teachers tend to enjoy more autonomy than they might within a language school. This is reflected in the fact that there appears to be less observation of sessions by CDs in universities than in language schools, where the hierarchical roles and responsibilities may be more clearly delineated. In one of the language schools surveyed CDs discuss the course they are directing on a daily basis with the Teacher Training Coordinator (Informant 5: 8), as well as being debriefed at the end of a course by the Coordinator and Senior Tutor (Informant 5: 30), suggesting a specialisation of roles and a level of institutional commitment that is unlikely to be found in a typical university department, where the CD's comments and conclusions will simply be presented in the form of a course report.

One factor which is bound to have an effect on the relations between staff and participants is whether or not participants are assessed. The possible effect of this on participant evaluations of a course has been discussed earlier, but the point is reiterated here because the knowledge that at certain points during a course the roles of assessor/assessed will be adopted may create a distance between staff and participants; but also, as a corollary, free those institutions for whom assessment is not an issue to experiment with other forms of non-threatening programme evaluation. This may explain the tendency for those institutions offering non-certificated open-entry programmes to personalise the evaluation

process through individual interviews and, in the case of at least two institutions, participant diaries.

There may also be relationships between institution type and programme length and between institution type and size of intake. For instance, the universities' main concern is the one-year programme, whereas the main business of the language schools is in much shorter programmes, normally of no more than one term in length (e.g. an intensive Cambridge/RSA Diploma) and often as short as two weeks in duration. Some of the latter may be run for as few as six participants. As far as evaluation is concerned, at least two effects spring from these differences: where the group of participants is small and the programme relatively long, evaluation can more easily be individualised (e.g. interviews in which feedback can be given and elicited); some language schools are able to do this even for short programmes. Where the intake is large and the programme long, it is more likely that there will be an institutional commitment to regular staff meetings and reporting procedures.

4.6.7 Attitudes to evaluation

4.6.7.1 Responses to the questionnaire

Table 4.7 summarises responses to Q.5 on the questionnaire:

Are you satisfied, in general, with the procedures currently used to evaluate your courses for teachers ?

Table 4.7: Tally of short responses to Q.5 on questionnaire

	M	B.Ed	Dip	OC	NC cl	NC op	TOTAL
YES	6	1	6	10	11	21	**55 (50%)**
YES/NO	4	1	1	5	1	2	**14 (13%)**
NO	10	3	2	3	-	2	**19 (17%)**
n.r.	1	-	1	1	13	6	**22 (20%)**
TOTAL	21	5	10	18	25	31	**110**

If we leave aside the disconcertingly high number of null responses (n.r.), these figures suggest a healthy level of self-criticism (30% selecting NO or YES/NO). However, when we examine the breakdown by institution type (Appendix 4.5), it becomes evident that levels of satisfaction are considerably higher in the language schools than in the

universities and colleges, where reservations (NO, YES/NO) are expressed about 31 of the 56 courses on which opinions are offered (equivalent to 56%).

Of the 38 respondents who completed questionnaires, 16 appear to be less than wholly satisfied with the evaluation procedures used within their institutions; only two of the ten respondents from language schools expressed any kind of dissatisfaction. A number of the constraints which affect evaluation have been detailed in the last section, but the main causes of concern expressed in the questionnaire returns can be summarised as (1) feedback from individual students (2) systematicity (3) post-course evaluation.

1. *feedback from individual students*: As we have seen, Respondent A regrets the fact that more time cannot be given on short programmes to individual tutorials. This is a point that is also taken up by Respondent L, who feels that more discussion is needed between students and staff and that ways need to be found of obtaining feedback from 'less forthcoming students'; Respondent M points out that 'many students do not participate in anonymous feedback' and this feedback therefore tends towards extremism because it derives from 'the vocal minority'. Informant 5/A describes two techniques for counteracting this tendency: dividing the class into small groups and putting the more vocal ones together; and getting participants to create or contribute to a poster display of feedback on specific issues (*cf* Morrow and Schocker 1993).

The need is also recognised for discussion among staff: Respondent N points out that on one of the programmes in which she is involved there are 'many different agendas' and yet 'we don't negotiate enough between ourselves as Ts (we have very different styles and expectations) and the students'.

2. *systematicity*: Respondent 1/B says that 'if problems arise ... we always wish we had evaluated more systematically'. See also Respondent P, below and the comments of the interviewees in 4.6.7.2, below.

3. *post-course evaluation*: Respondent I, as we have already seen, indicates that his institution has not yet developed post-course evaluation procedures; Respondent O also notes 'no follow up'. Respondent 1/B recognises the advantages of post-course evaluation

'when it is possible'; Respondent P admits that post-course evaluation is 'not as comprehensive and systematic as it might be'; and Respondent Q comments on a course leading to the Cambridge/RSA Dip. TEFLA: 'I think post-course evaluation might be useful ... I hadn't really considered it before as DTEFLA is an exam-based course'.

The fullest response to this question among the questionnaire respondents comes from Respondent H, who sees deficiencies at all points in the systems employed within his institution. He feels that there could be better pre-course analysis of needs and wishes through questionnaires and observation; more use of interviews as part of in-course monitoring and end-of-course evaluation; and better post-course evaluation in the form of analysis of the relevance of course modules to actual professional needs or 'tracer studies to assess the economic impact of participants'. As noted earlier, he also feels the need for more thorough analysis of the information that is gathered.

4.6.7.2 The interviews

Interviewees also admit to reservations. Indeed, as they talk they seem to become more aware of these. Towards the end of her interview, Informant 4 concludes: 'I think we need to reflect a lot harder on what we're doing and try and make it more systematic' (20). The interview with Informant 1/B is particularly interesting in this respect:

> ... during the course I'm not looking for them to say we got our money's worth or we feel the time we're spending here is valuable and worthwhile. **As I say that I realise that perhaps I should be**. At the end of the course at the end of the course I really want the reassurance that they feel the course has been worthwhile and they've gained a lot from it [*Hm*] and it's important [...] to know the extent to which they have felt that and if they haven't to know why and what we could have done differently (pause) **I'm realising how little I've thought thought it through as I talk** (pause) and (pause) I think that if at the end of a course a group is really happy I kind of think it doesn't really matter what they say [...] we've done our job and that's it and I don't really find out what it was that made them happy [...] if they've been dissatisfied then I'm much more concerned.
>
> (29, emphasis added)

At a later stage in the interview, during a discussion of questionnaires, he is asked if any quantitative data is collected on the basis of which decisions can be made:

> Impossible to process (30).

[...]

Are they open-ended ?

Hm. Totally ! (32)

He is then asked if he is happy with the open-ended format, if it suits his purposes:

> I don't know what the answer is. **Part of me is sitting here feeling a bit embarrassed and feeling feeling I think I've been rather smug about what we do. Part of me is sitting here and thinking (pause) I mean I haven't really thought through the question so I don't have an answer to it**. My instinct part of my instinct is to say that er quantitative data (pause) is not so useful (pause) but maybe it is maybe we should...
>
> (33, emphasis added)

See also 4.6.8.4.

Informant 3 appears to have few doubts about the general approach within his institution during and immediately after a course:

> I think I'm reasonably happy with them as they happen during the course. I mean, I think that the system of collecting information, I think the kind of information we ask for, the frequency with which we ask for it, the what we do with it - because it's all summarised and discussed with the participants and you know we make changes according to how they and us feel about it all erm - and then the final evaluation and then the other ones are all used to make the report on the course and that's fed into the next course or other similar courses. So I think that that procedure is is is all right really because it's it's I don't think we're overdoing it I don't think we're underdoing it we're getting reasonable information erm and so on...
>
> (7)

However, he continues:

> ... but I think that in terms of any real idea of the value of the courses to the participants in their own teaching situations I think about that aspect of it we are far more in the dark about and really lack adequate information. And erm it's it's vital to have it and so we go on operating on hunches and guesses and assumptions perhaps quite unwisely erm I'm sure very unwisely in many cases because of the lack of information of that kind. Erm and in fact I think I would go so far as to say that it's impossible to get it. I think that's the point. And that in a way makes one could make one question the value of courses of this kind that are so removed from the erm context and which it is so difficult to get meaningful information about.
>
> (7)

As Informants 5/A and 3 acknowledge, the real problem lies with open-entry programmes:

> With us a course leaves on a Friday afternoon and that's that's the end that's [*Hm*] that's the last we see of them. No it's not very satisfactory we haven't a clue whether erm what they have learnt is is going to be used if it is at all useful at all erm whether it was even interesting for them.
>
> (Informant 5/A: 46)

> with the open-entry courses it's really a problem... and I I think that that's a serious flaw in our operation. Erm erm and and I I don't know a remedy to it. The only remedy I can think of is the one that would talk me out of a job (laughs) unfortunately (laughs). In other words you know a great deal more of the training should be based in [*Yeah*] in their own country in their own school and [*Yeah*] what have you.
>
> (Informant 3: 7)

Informant 3 is perhaps unduly pessimistic. If the objectives of non-certificated open-entry courses relate to what can realistically be achieved within the course then information on post-course effects may be interesting or reassuring but not essential.

4.6.7.3 Interpreting the responses

The implications of this analysis of the reasons for dissatisfaction are taken up in section 4.8.2, *Towards appropriate evaluation*. There remains the question of the positive responses and the value that we attach to them.

A positive response may be interpreted in two ways:

1. *I am satisfied that we evaluate thoroughly.*
2. *I am satisfied that we do as much as is feasible and appropriate, given the nature of the programme and the resources available.*

An impression of the likely validity of the opinion expressed in (1) can be obtained by reference to the range of evaluation procedures used. The validity of (2) is less amenable to objective analysis since it is dependent on a subjective view of what is feasible and appropriate; nevertheless, comparisons may be made with other institutions offering programmes of the same type.

Quantitative analysis of responses to sections 2-4 of the questionnaire (in-course monitoring, end-of-course evaluation, post-course evaluation) in respect of selected programme types reveals the following:

Master's and B.Ed programmes

- range of procedures used: 6-19 (maximum possible 17, plus additions under 'other')
- mean: 11.7

The respondent from a university using 8 procedures expressed him/herself satisfied with the procedures used; the respondent listing 19 procedures professed herself dissatisfied (no post-course evaluation).

Cambridge/RSA Dip TEFLA and other certificated programmes

- range of procedures used: 5-16
- mean: 9.3

Four institutions list 6 procedures and two only 5 procedures. Of the latter, one - claiming to be satisfied with these procedures - seemed undecided as to whether participant results were taken into account in evaluating the course (a question mark was inserted next to the answer box), and the other gave a null response to the question concerning satisfaction and apparently sets no store by participant results. All six courses at this end of the evaluation spectrum were validated by examination boards, which implies that some institutions are content to rely on the systems established by the examination boards.

Non-certificated closed-group programmes

- range of procedures used: 3-17
- mean: 9.9

Interestingly, the top and bottom of the range are represented by the same institution (a key difference being programme length); the respondent did not answer question 5 for the minimally evaluated programme but pointedly answered 'YES - for this course' in respect of the other.

Non-certificated open-entry programmes

- range of procedures used: 0-8
- mean: 6.7

As we have already seen, one institution does no formal evaluation of its 5-day programme; the respondent from an institution using 3 procedures also claimed to be satisfied with this. The majority of the programmes described were of 2-3 weeks' duration.

The general issue here is that of the criteria being used to assess the systems of evaluation operating within individual institutions. The comparative figures may be suggestive of differences in standards of judgement, but without some consensus on the criteria for assessment and a good deal more information about how specific procedures are used, it is difficult to draw any firm conclusions about the validity of the opinions expressed in the questionnaire returns.

Another source of data is available, however, in the form of the interview data, and this allows for what was referred to as 'triangulation of ... method' in Chapter 3 - that is, the comparison of the products of two methods of data collection. Unfortunately, comparative data is available from only three sources, Informants 1, 4 and 5; nevertheless, it is relevant to note that in their completed questionnaire returns neither 1 nor 4 had given any indication of the reservations expressed in the interview extracts (above). We may therefore infer that actual levels of satisfaction would, on further investigation, prove to be lower than those shown in Table 4.7 at the beginning of this section.

4.6.8 Further issues

In this section we consider a number of other issues touched on in the interviews. These issues are:

- the reliability of participant opinions as evaluation data;
- the nature of intended effects and criteria for assessing these;
- the use made by institutions of evaluation data;
- the extent to which evaluation procedures are modified over time.

4.6.8.1 Participant opinions as evaluation data

Mixed opinions are expressed about the validity of the opinions expressed by course participants. We have already seen that some questionnaire respondents feel doubtful about the representativeness of the feedback they receive and are also conscious of the kind of end-of-course euphoria which can bathe a course in a rosy glow. There is, nevertheless, a readiness to trust participant feedback because of the nature of the relationship that has been established between participants and tutors. Thus Informant 1/B wonders whether participants on an assessed course 'feel that if there is any doubt over their grade they may be prejudiced by criticising the course', but concludes nevertheless: 'Perhaps this is utterly naive ... I think that the climate of openness and honesty and trust is such ... I believe this ... in this place that by and large people feel comfortable about giving honest feedback' (13). Since in some cases there are no other measures of the success of a course, this position may be understandable, but it is not very sound. Unless evaluation is anonymous, the expressed opinions of participants on award bearing courses are likely to be influenced not simply by anxiety about their grades but also about reports (to sponsors) or job

references; a positive relationship between tutors and participants can also distort results if a sense of loyalty or concern for tutors' 'face' influences what is said and withheld. In general, then, confidence in participant feedback has more validity when it relates to post-course comments (as argued by Informant 2) and/or has some corroboration from independent sources.

4.6.8.2 Intended effects and criteria for assessing these

Programme objectives vary and it follows that the instruments used to measure them should vary accordingly. If the focus has been on developing participants' linguistic knowledge, then one appropriate way of assessing the attainment of this objective would be an end-of-course test, but if a related objective has been to develop participants' confidence in using English, say, as the medium of instruction, this can be evaluated either by an analysis of attitude-change or (more reliably) through classroom observation. Given that our central concern is INSET courses and that some of the desired effects of such courses will only be observable as a result of classroom observation, this is a further argument for some form of systematic follow up of participants.

Informant 4, whose institution does engage in a certain amount of follow-up activity, was asked how she assessed programme effects and whether there was any difficulty in discerning these. The answer suggests that this aspect of the evaluation process is less than systematic:

> *How standardised is the observation that you carry out of the teachers who've been on a course here ? So you want to see what effects the course has had, you go into a number of classrooms* [Yeah], *what are you looking for ? Are you looking for the same thing in every classroom ? ... How do you evaluate ?*

How do I evaluate ? Um just um I suppose it's interesting to go into the classroom and see whether anything any of the techniques that we suggested are employed. Sometimes you go in and the teacher's just performing in exactly the same way as they always have done [*Hm hm*] a lot of mother tongue a lot of things happening in French and then you go into another classroom and you'll see somebody actually developing some of the ideas that were developed in [PLACE] and you might go into another classroom and see somebody's who's moved on and actually produced something [*Hm*] in France as it were [*Hm*] So I think I'm interested to see whether the course here has had any impact ... We found the interesting thing is that although French teachers within their own school don't collaborate very well one of the products of this erm this networking thing is that quite a few of them get together [*Hm hm*] once every two months [*Hm*] for a kind of reunion and [...] exchange material [*Right*] ...

How can you be sure that what you see is different from what you would have seen if you'd been into that classroom before the course?

We can't ...

You don't get a chance to do a before and after ?

We don't ... only on one or two occasions. Sometimes we get a chance to see teachers who we taught like three or four years [*Hm*] before. That's that's quite interesting. And I think on the odd occasion we've been able to observe one or two participants before but that's not normally the case.... I wish we could but it's really the financial restrictions.

(8-10)

She goes on to explain that the cost of the visits to France is borne by the institution and that, as far as she knows, the other centres operating this programme do not carry out visits of this kind, although all the centres are obliged to attend an annual meeting in France. She is then asked whether she can trust what she sees.

How do you know that what you see when you go into the classroom if it does appear to draw on what's been done on the course isn't just being put on for your benefit as it were [Yeah] *as opposed to being you know something that's been fully integrated into* [Yeah] *this teacher's* [I know] *repertoire?*

I think often the reactions from the pupils tell you (laughs) tell you quite a lot as to whether they're used to this or whether it's something very novel. Ah again discussion with the with the teacher erm [*Hm*] usually. They tend to be because we it's quite a close group usually after the three weeks everybody's bonded quite well [*Hm*] and there is quite a degree of frankness [*Hm*] and er talking through the problems of the classroom situation [*Hm*]. Because obviously schools differ [*Hm*] and in some schools it's easier to implement new things [*Hm hm*] than others.

(11)

A similarly impressionistic approach is described by Informant 5/A (34). Judgemental though they may seem in print, the interviewer's *Hm*s were intended to be encouraging:

How would you know that what they'd be doing when you observed them was different from [Dunno dunno] *what they'd be doing* [haven't a clue] *before ?*

I just wouldn't know at all [*Hm*] would just have to go by their word [*Hm*]. I mean you can tell up to a point as to whether the students are responding in a surprised way (laughs)

However, there were benefits in merely going to a country to look at conditions in schools:

I think the main thing that we learned from it was just sort of what the level of the trainees really was. I mean what the teaching situation really was, what was actually possible ... it's not until you actually get into the schools and just see a forty-minute period [*Yeah*] what you can achieve in a forty-minute period ...

(34)

She goes on to describe a more focussed visit for a week-end seminar with forty former trainees who had followed a course in language development and basic methodology. Although evaluation was not the main purpose of the visit, she discovered that only one of the forty had subsequently tried to implement anything new. In this case, action was taken as a result of the evaluation:

Was he doing it as a result of the course ?

Yes. [*Ah ha*] And he was very enthusiastic about English and erm he just loved it [*Hm*]. But the others didn't (laughs) and so they were taking the easy option [*Hm*] ... I mean they had a very very heavy teaching programme diabolical books so it was hardly surprising (laughs)

Did it force you to think again about what you did ?

Well, what we did we actually changed the focus of the the course erm to be predominantly language development.

(40-41)

She also describes what she plainly feels to be the ideal situation (emphasis added):

... we had a very big [NATIONALITY] project ... trainees coming to us first for language then they went back to their country for initial training in methodology then back to us for a three-month brush-up [*Hm*] on ... you know this is how they do it in England to get enthused (?) about it, back to the schools where they were monitored in the schools and we went out and worked with them in their schools for ... four weeks at a time [*Hm*] and saw them teaching and had workshops with them ... every week ... And this went on for several years ... Post-course evaluation wasn't just post-course evaluation erm it was preparing for the next stage of another course [*Hm*] so yes **that was real evaluation because you were seeing the results of one lot of input and how it fed into another rather than something that was static**.

(45)

Informant 2 has fewer expectations in terms of short-term effects on those who have followed the Master's programme at his institution:

It sounds to me from what you've said that sponsors are relying - as you rely - on participants' perceptions of the value of what they've done [Yes] *rather than any let's say changes in behaviour in classrooms, which is presumably what you're really part of your ultimate aim is to change the way people think about what they're doing and therefore change what they're doing and the way in which they do it.*

I think you have to bear in mind erm that we work more with trainers than with teachers and erm that the er that the kind of measure of change, the behaviour, has more to do with trainer change than with teacher change. Of course, there's a knock-on effect ...

You mentioned the knock-on effect.... You're dealing mainly with trainers who deal with teachers who in turn deal with learners. Isn't this one of the problems for people who're trying to evaluate effects, that it's terribly difficult to get to the the sort of ultimate effects of seeing changes in learners as a result of what [Very very difficult] *has been done with teachers or trainers ?*

> Very difficult because I mean six months would be much too early along the line to evaluate that and the knock-on effect er might take years.
>
> (18-19)

There is undoubtedly an issue to do with the transparency of effects and the time needed for effects to manifest themselves. If the resources committed to post-course evaluation are to be properly exploited, however, careful thought is needed as to what effects are sought and how these might be evaluated.

4.6.8.3 Use made of evaluation data

Records are an essential input to evaluation (see, e.g. Informant 1/B: 4, Informant 2: 26). Although various purposes for evaluation are distinguished (e.g. by Informant 1: 1), the emphasis in the interviews is on evaluation as an input to fine-tuning or more radical course development (Informant 2: 14; Informant 4: 12; Informant 5: 41, 51). Other uses are, however, mentioned. For instance, Informant 5 (25) explains that individuals might be moved to another group during a course and Informant 2 talks about the use made of negative findings concerning a tutor:

> *What kind of internal discussion takes place as far as the course report is concerned ?*
>
> ... It very much depends on whether the course is part of a pattern that's meant to be repeated or whether it's a one-off. Erm if it's a one-off what we tend to look at in reports are those issues that might have staff development er consequences like erm if a particular tutor has obviously not sussed out what the course is about or what's required and then we pick that up and work with the tutor concerned, If it's a course that's likely to be repeated erm we're probably going to study [more areas?] talk about it and erm and try to ensure that whatever mistakes we made on that course would not be repeated. And that of course also has a staff development dimension in that erm the people concerned the tutors concerned are erm made conscious of things they have to work on in their own training style or whatever.
>
> (27)

In one institution, standard questionnaires required by the bureaucracy appear to disappear into a black hole (Informant 5/A: 26). Programme personnel might be forgiven for taking evaluation less than seriously if nothing appears to be done with the results.

4.6.8.4 Modifications to evaluation procedures over time

If institutions are serious about evaluating their programmes, one might also expect that they would be concerned to evaluate the instruments they use for this purpose, since an evaluation is presumably only as reliable and sensitive as the instruments used.

Such evidence as is available from the interviews suggests that there may be a tendency towards conservatism on this score. Informant 3, for instance, is pressed a little on the evaluation of short closed-group courses:

> *because they've been running for some time with a similar clientele* [Hm] *you don't feel the need to revise the evaluation procedures very much ... ?*
>
> Yeah, well, I think that I should add that as far as these courses go which er we've been getting for some time you know with similar personnel another factor is that the same staff have been involved with them [*Hm*] for some time as well and so erm in in I think that makes us, yeah, that definitely makes us feel that although there's always something to learn and there are always things you should be watching out for erm and you know there's always something you want to change at the end of the day or before the end of the day it's not quite the same as not at all the same as really as dealing with a course which is much more novel in all of in any of these respects. So I think that wisely or otherwise that is the the view that we've adopted about those courses.
>
> (4)

Informant 4 is a little more forthright about her reservations:

> I think possibly procedures are just used because that's what they did that's what was done last time rather than thinking [*Hm*] well perhaps we could just concentrate on one one form of evaluation like give diaries a go or [*Hm*] because sometimes I think the students end up feeling can end up feeling overevaluated.
>
> (20)

Asked whether evaluation procedures have changed in his institution over time, Informant 1/B indicates that more evaluation now takes place. The impression given, however, is that the increase in evaluation has simply yielded more data of the same kind:

> I've I've been fairly kind of satisfied about the way we do feedback without really or the way we evaluate courses without really thinking about it my instinct you see and this is the kind of smug side of course my instinct is that that we are (pause) that we are in touch with how people feel about their course [*Hm*] of course it doesn't that's not the same as being in touch with what they're learning during the course.
>
> (33)

When a system is working well, there is some justification for doing 'what was done last time', as Informant 4 put it, although some might argue that there is a danger in such complacency[13]. The evidence from the survey is that systems within institutions are not working as well as they might and there is some awareness of this. In such a situation, there is an even stronger case for reviewing the system.

4.7 EVALUATION OF SURVEY INSTRUMENTS AND PROCEDURES

4.7.1 Survey design and administration

The decision to collect data by means of interviews as well as through questionnaires seems to have been amply justified since the interviews served not only to supplement, as anticipated, the questionnaire data supplied by individual institutions, but also extended the set of institutions included. It is possible that the return-rate could have been further improved by using some of the techniques referred to in the literature (e.g. follow-up letters, telephone calls, postcards). However, if there is any validity in the analysis (in 4.4.6) of the null responses, this would not necessarily have made a substantial difference.

The actual coverage, although smaller and more limited than desirable - particularly in respect of the state colleges - may be sufficiently representative for generalisable conclusions to be drawn (the issue of possible bias having been dealt with under 4.4.7). This assumes, however, that the responses given can be taken to reflect the reality and that terms such as 'regular', as in 'regular meetings' and 'discussion ... at regular intervals', mean the same to everyone. As the last subsection implied and the next makes abundantly clear, it may be wise to reserve judgement on this last point.

4.7.2 Evaluation of the questionnaire

Although in general the questionnaire seems to have elicited the kinds of information it was designed to, there must be some doubt about the interpretation of certain terms by particular respondents - and this may indicate a lack of clarity in the formulation of concepts (or, in fact, in the way in which these terms are generally used within the profession).

One respondent queried (spoken communication) the concept of 'programme evaluation': was this, he asked, evaluation of courses or evaluation of participants? The use in the Courses Checklist of the term 'open-enrolment' also caused difficulty ('As you will see, none of our courses have open-enrolment, and for many of them there is a very exacting selection procedure' - from a letter accompanying a questionnaire return). One interviewee used the term 'closed group' as opposed to 'open entry', and since this distinction between 'open' and 'closed' is clearer than that between 'open' and 'contract' (the latter term being the one used in the Courses Checklist), these terms have been adopted in this chapter.

A more problematic concept appears to be that of 'in-service'. Two respondents from universities included without comment details of MPhil and PhD programmes in their returns; a third university respondent commented on the post-course section of the questionnaire 'irrelevant for our course; we are not engaged in teacher training' (original emphasis).

Asked about the term 'in-service', one of the interviewees, who had been describing the programmes offered by his institution, had this to say:

> if you see in-service as meaning having already taught erm yeah for well a a reasonable length of time then I would have thought that there was no question that these are these are in-service.
>
> *Is that how you see it ?*

> Ah well, now this is a good question. You make me realise that I've not really thought about the meaning of the term. Erm no, I would in fact say that it really implies that people have had some kind of initial training as well as prior experience and indeed that that well strictly speaking I suppose you could say that the experience is on-going and er and but yeah that would apply I think. It needn't be in the same job it can be between one job and another. Erm, so as as we talk it makes me think erm whether I do have an adequate concept of the word in-service, but yes I think it should include both at least both of those things: a certain number of years of having taught with an intention of going on to continue to do so but also already got some had some kind of meaningful initial training and this is in-service training ... something that builds on those initial experiences and qualifications.
>
> (Informant 3: 2)

The inclusion of prior training as one of the two givens makes this a somewhat ideal definition, at least as far as many TESOL teachers are concerned. Perraton's (1993) distinction between pre-service and in-service *initial training* on the one hand and *continuing education* on the other is a helpful way of separating out the training and experience variables.

What prompted Respondent Y to include his Cambridge/RSA Certificate courses in the return was the fact that some participants on these courses already have experience of teaching. Although this type of course was excluded from the data, the issue remains of whether Certificate courses in institutions which have a similar selection policy (i.e. favouring applicants with prior teaching experience) should not have been included in the survey. The distinction currently being discussed within UCLES between a pre-service certificate and an in-service certificate (the latter being a qualification for those with experience but without an initial teaching qualification) might go some way to resolving the problem of labels - and inform syllabuses - at this end of the training/education spectrum.

Two other terms in the questionnaire prompted unexpected responses. The first was the inappropriate use of the term 'validated' in 'validated by the external examiner', about which one careful respondent remarked: 'The degree is internally <u>validated</u> but there is external moderation by external examiners' (original emphasis). The other term was 'post-course evaluation', as distinct from 'end-of-course evaluation'. Under the former, two respondents included under the *Other* heading 'external examiner's report', an evaluation procedure which, despite being temporally later than the end of the course, is conceptually

closer to end-of-course evaluation than to post-course evaluation (which had after all been glossed in the questionnaire as 'e.g. 3 months after course').

It would seem that in reference to teacher education and evaluation one cannot take for granted a shared set of terms and concepts. Nor, indeed, can one take particular practices for granted - for instance, that data which could be used as input to programme evaluation is actually used for this purpose. Reference has already been made to the use of pre-course information for this purpose. Respondent Z, in relation to end-of-course evaluation, similarly queried the use of tests/exams/assignments as input to programme evaluation.

Reference was made above to the 'other' or empty slot option. In principle, this ought to catch the odd unpredictable response rather than serving as a catch-all category. However, as can be seen from Appendix 4.4, the responses under 'Other' were more numerous and varied than might have been expected. The point at issue here is not only the accuracy or otherwise of the predictions, which were perhaps less comprehensive than they might have been, but the fact that, so far as the additional options supplied by respondents are concerned, the results can only be suggestive - other institutions may use these procedures but not have thought to mention them. With hindsight, the following options, all of which were mentioned by at least three respondents, should have been included:

1. under **pre-course evaluation**:
    - references
    - letters of application/application forms
2. under **in-course monitoring**
    - student representatives/staff-student commmittees
3. under **end-of-course evaluation**
    - External Examiner's report
    - exam board (for certain types of award bearing course)
    - Course Director's report
4. under **post-course evaluation**
    - follow-up visit to monitor effects

4.7.3 Evaluation of the interviews

The value of the interviews in eliciting information and opinions that could not easily have been collected by other means is amply demonstrated by previous sections.

A secondary and unexpected outcome was the effect on respondents.

One respondent (P) noted at the bottom of the questionnaire: 'Filling in this questionnaire has made me think about evaluation of the course more seriously' and the interview prompted Informant 3 to reconceptualise what his institution has been doing:

> because that's such a long programme that's been going on for about five or six groups now, there've been a lot of visits backwards and forwards .... And I suppose yeah I mean I hadn't really thought of using the word evaluation for it erm but a great deal of it is really that when you get down to it.
>
> (5)

Informant 2, asked if he would want to introduce forms of evaluation other than those currently being used, says:

> Perhaps one of the things we could benefit by would be ... something you alluded to earlier on [in a question] during a course for a colleague who's not teaching during that course to come in and could find out what's going on in the course and feed back to course tutors. I think that's one thing that I I would possibly do that would be useful to the growth of a course while it's going on.
>
> (34)

The effect on Informant 1 while the interview was in progress has already been illustrated (4.6.7.2); this was confirmed when he returned the transcript:

> You may be interested to learn that I have been thinking and talking about a number of things arising directly from our conversation three weeks ago - in the main, how do we evaluate what they have learnt ?, and the point about quantitative data (written personal communication).

4.8 CONCLUSIONS

4.8.1 General conclusions

It seems to be normal in reporting surveys to adopt an apologetic tone in reference to the return-rate (and 4.7.1 makes a gesture in this direction). In this case, however, in which institutions were being asked to supply information of a potentially 'sensitive' nature, it might be more appropriate to celebrate the fact that so many institutions were prepared to be open about their practices in the interests of the research study and ultimately the profession at large. This openness and the constraints under which all or most institutions operate are articulated by Informant 2:

> We don't do the same kind of follow-up on three-week courses ... or three-month courses that we have. Erm on those courses we do an end-of-course questionnaire and we take that at face value.
>
> *Does that mean you're satisfied with it ?*
>
> No. What it means is that there's a limit to the kind of endeavour that we put into follow-up and evaluation and when we have a three-month programme of one sort or another we we recognise that people are just just about with us for those three months but they're still very much focussed on what they've got to do when they go back and erm that imposes limitations, say. We get the data that we can then there and then [*Hm*] I I think that if we were honest about I mean the problem is erm always with evaluation erm the more sophisticated you make your evaluation instrument the more demands it places on you as a lecturer as a team of lecturers to to do something about it. And erm there's just simply a limit to time and er a limit to er how many different areas of focus you can cope with at any one time and so on.
>
> (15-16)

He concludes in similar vein:

> All our efforts at evaluation have been definitely influenced by manpower problems and time problems [*Hm hm*]. We do as much as we can within the limits that we've got but we don't ... What I would find counterproductive would be if the data we were getting from the evaluation erm was more than the tutors' team could cope with and sometimes I think we get very near that.
>
> (34)

The findings of the survey are revealing in a number of ways. Although there is evidence of concern for evaluation in general terms, it seems to be the case that this concern is not always well informed or well focussed: data that is necessary for systematic evaluation (e.g. concerning post-course effects) is typically not collected; data that might be of potential value in evaluation is not utilised or is reportedly not thoroughly analysed; and data collection instruments that might be useful are not employed. There appears to be some awareness of these deficiencies but it does not extend to all institutions and, in general, a fairly pragmatic attitude prevails regarding the evaluation of short courses. However, if the shifts in attitude which seem to have been stimulated in some of the informants by the survey are anything to go by, further discussion of evaluation within the profession might prove beneficial.

4.8.2 Towards appropriate evaluation

The findings of the survey give rise to a number of questions.

ñ *If the data that is normally available at the outset of a course can make a contribution to evaluation, why is it not used for this purpose ?*

ñ *Should institutions make more of an effort to collect baseline data specifically for the purposes of evaluation ?*

The value of before and after comparisons has been argued at earlier points in the thesis.There is no evidence from the survey to suggest that any systematic efforts are made to carry out comparisons relating the end-performance (or post-course performance) of participants with their starting-points. This issue is taken up in Chapter 5.

ñ *What forms of evaluation are desirable and feasible on short courses (maximum three weeks) ?*

ñ *What should the criteria be for such evaluation ?*

There is evidence from the survey that course length is an issue and although the vast majority of institutions do carry out some form of evaluation of short courses, this may be very limited, and typically makes use of an end-of-course questionnaire and whole-group discussion (as a monitoring device or on the final day). A principled basis for any decision

as to which procedures to use ought presumably to stem from a statement of the objectives of the course and of the objectives of and criteria for the evaluation. If the primary objective of the course is to 'refresh' jaded teachers, then this will have certain logical consequences for decisions concerning evaluation: e.g. the primary criterion for evaluation will be whether (and if so, the extent to which) participants feel 'refreshed', however this is defined; and the most appropriate way of establishing this will be to ask them, at the end of the course, how they now feel, using for this purpose a questionnaire, whole-group discussion or individual interview, or some combination of these.These questions are discussed further in Chapters 5 and 6.

ñ *What forms of post-course evaluation are desirable and feasible ?*

For courses which are designed to produce effects which are observable in classroom performance in the short term, the most appropriate form of evaluation would obviously be to observe participants some months on in their own teaching contexts. Since this is built into some projects and closed-group programmes, it is clearly feasible if the resources are available. Where this is not possible, a report from someone in the field such as a sponsor may be helpful; indeed, a sponsor's report may provide a useful complement to a course tutor's own observations. Where classroom observation is either impossible or not felt to be particularly appropriate, a post-course questionnaire is an obvious if minimal (and not always successful) form of follow-up. Some thought is needed on the presentation of this and its timing. An alternative to asking participants to complete a questionnaire is to ask them to report on progress in achieving an action-plan of the sort referred to in Chapter 3. Since the plan has been formulated by the individual participant, it is more narrowly focussed on his/her concerns than a questionnaire might be and this might result in a higher response-rate; the limitation of any response is that it may hold good only for that participant. The feasibility and value of using action plans for programme evaluation is explored in Chapter 8.

ñ *How far can we rely on what participants tell us ?*

ñ *Which forms of evaluation are most reliable?*

The problem of getting feedback from the whole of a participant group and of getting honest feedback has been discussed in earlier sections. Individual interviews may serve the first purpose, as might an in-course anonymous questionnaire, but they may still fall short of that level of honesty that might be desirable. Alternatives include participant journals as an in-course monitoring tool and the use of someone outside the course team to observe or elicit participant comments. The utility of both these procedures is explored in Chapter 7.

- *Do we sometimes do too much evaluation ?*

There are, of course, limits to the amount and types of evaluation that can be carried out. As we have seen in this chapter, there will be constraints imposed by the time that can be made available within the course for data collection and analysis; there may be a problem of participant accessibility, especially after a course has ended; and there may also be a point (amount, point in time) beyond which participants may not be willing to cooperate. These are, in a sense, the limits of feasibility.

There is also the question of utility and frugality. Do the measures used elicit useful information economically ? Do we over-evaluate (e.g. collect data that we do not analyse thoroughly ? stick to established routines even when these are producing 'known' information ?)? These questions are taken up in Chapter 6.

4.8.3 Course-sensitive evaluation

Although it is possible to identify evaluation procedures which are used on all the programme types surveyed, others (indeed, clusters of procedure) are used only on programmes of a specific type. Some of these procedures are imposed by examination boards or agreed by negotiation with a sponsor; others form part of the moderation/ validation norm across institutions or within a particular institution. This leads to the rather unsurprising conclusion that up to a point the decisions concerning the evaluation procedures to be used will be specific to programme type. But perhaps we can go beyond this. Even when a course is well established in the sense that it has a set of agreed objectives, a timetable and materials, the participants on a particular course within that programme may vary in number or origin or needs or wants significantly from those on previous courses; moreover, as was noted earlier, the world of ideas, awareness and

teaching practices outside the institution where the programme is held does not stand still. These considerations argue not simply for evaluation that is specific to a type of programme but to that particular programme and each occasion that the programme runs; and this we might term course-sensitive evaluation. The concept of course-sensitive evaluation underlies the Case Studies in the next four chapters, but is most fully developed in Chapters 7 and 8.

CHAPTER 5

CASE STUDY 1

A SEMINAR FOR ML TEACHERS

5.1 INTRODUCTION

Access to information on participants' expectations, knowledge, attitudes or practices prior to the commencement of a course is not always possible, communication difficulties or late enrolments being two reasons for this. When it can be obtained, such information can be a significant input to the modification of an existing programme, as noted in 2.6.2; it can also serve as a yardstick, or baseline, for course evaluation. Answers to section 1 of the survey reported in Chapter 4 (see 4.6.3.1) suggest that in some cases institutions make no attempt to obtain pre-course information and in others that even when the information is available it is either not used at all or its value is perceived only in relation to course design.

This chapter presents a small-scale study in which a systematic attempt was made to gather information on participants prior to and following a single seminar within a continuing programme. The data-collection methods used (questionnaire and interview) were those most commonly used by respondents to the survey.

The research questions which informed the study were as follows:

1. *in what ways can information obtained prior to the commencement of a programme (or course) inform evaluation ?*

2. *how suitable are questionnaires and interviews as means of collecting the required information ?*

5.2 THE STUDY

5.2.1 Introduction

Staff seminars lasting 1½-2 hours are a regular feature of the professional development programme at the Institute for Applied Language Studies (IALS), University of Edinburgh. This evaluation study relates to a staff seminar on the use of the overhead projector (OHP) for teachers of modern languages (henceforth ML) which I conducted in 1993. The seminar topic had been specified by ML staff. At the time of the study, 35 teachers were involved in ML teaching, the vast majority of these being part-time.

5.2.2 Evaluation questions

In evaluating the seminar, I had two objectives: one related to the seminar *per se* and one relating to the programme as a whole. These were to establish:

1. if the seminar was effective

2. what lessons could be drawn from the evaluation which would be of value for future seminars.

5.2.3 Stages in the study

1. It was hypothesised that, since the seminar had been requested, there was reason to suppose that a proportion of ML staff were not currently using the OHP, possibly for attitudinal reasons, and that those who were might be looking for new ideas.

2. A questionnaire (Appendix 5.1) was therefore prepared and distributed to all ML teachers in advance of the seminar. The questionnaire was designed to elicit information on the following: (a) extent and frequency of previous OHP use (b) purpose of use and (c) teachers' attitudes to the OHP. It also served to notify teachers that a seminar would be held, and gave details of topic, date and time. In order to facilitate pre- and post-comparisons, respondents were asked to give their names.

3. Of the 35 questionnaires distributed, 17 were returned before the seminar. These formed a basis for decisions on the content and emphases of the seminar. More importantly for the case study, they supported the working hypotheses noted in (1), above, and provided a baseline against which to assess changes in behaviour and attitude as a result of the seminar.

4. Ten teachers were present at the beginning of the seminar, three having earlier apologised for not being able to attend owing to teaching commitments elsewhere. Of the ten who attended, two left early. Details of the seminar participants, two of whom had not returned the questionnaire, are given below under **Seminar participants**. No attempt was made to elicit participant evaluations at the end of the seminar. This was a deliberate feature of the research plan. Such seminars are not normally evaluated, and to avoid the Hawthorne effect (Schwartzman 1993) I did not wish to alert participants to the fact that they would subsequently be monitored.

5. Three weeks after the seminar, structured interviews were held with each of the eight teachers who had stayed for the whole of the seminar. The basis for the interview was a questionnaire (Appendix 5.2) which was shown to interviewees but completed by the researcher. The interviews sought to establish the effectiveness of the seminar as reflected in changes in the frequency and purpose of OHP use and in participants' attitudes to the OHP as a teaching/learning aid. The two seminar participants who had not returned the first questionaire completed it at this time.

6. To investigate longer-term effects, certain subjects were interviewed again the following term. The decision to include this stage, which had not been part of the original research plan, was taken following the first interviews.

5.2.4 Seminar participants

The characteristics of the eight teachers who attended the whole of the seminar are summarised below:

- Languages taught: German 3; Italian 2; Spanish 2; French 1.

- Language teaching experience (years): 2-20 plus.
- Position held: full-time 1; three-quarters contract 2; hourly-paid 5.

These characteristics suggest that they were a reasonably representative sample of the ML staff. However, a comparison of responses to the pre-seminar questionnaire by attenders and non-attenders suggests that the former - the subjects of the study - were not wholly representative either in terms of OHP use or attitudes to the OHP (see Table 5.1, below). Since the purpose of the study is not to make generalisations about the ML teaching staff within IALS, these differences are simply noted and not explored further.

5.2.5 Seminar content

The seminar plan consisted of four stages, with rough timings:

1. elicitation of prior experience (on the assumption - based on the questionnaire returns - that there would be staff present with experience of using the OHP, and that this experience could be tapped for the benefit of all): 10 minutes
2. illustrated survey of techniques of use and purposes of use, concentrating in the latter case on uses which - judging by the pre-seminar questionnaire and what came out of Stage 1 - were not widely known: 50 minutes
3. hints on transparency production, electronic and otherwise: 15 minutes
4. transparency production (transparencies and pens supplied) and/or practice in handling the hardware: 15 minutes.

In the event, Stage 1 produced very little because with two exceptions participants had little or no prior experience of using the OHP. Partly for this reason, Stage 2 took up more time than anticipated, and no time was available for Stage 4. Participants' reactions are referred to in section 5.2.7, below.

5.2.6 Results and discussion

5.2.6.1 Pre-seminar questionnaire

Taking the two late returns into account, the questionnaire was eventually answered by just over half the potential number of respondents (19/35). For the purposes of the case-study

(and the seminar) responses were only required from seminar participants and no attempt was made to follow up non-respondents. One may speculate that those who did not respond knew that they would be unable to attend or had no interest in the topic (either because they were frequent OHP users or felt no need for the OHP, explanations supported by six non-attenders who did respond - see in Table 5.1, below, response (d) to Question 1a and response (a) to Question 1b).

Tables 5.1 to 5.3 present the results of the questionnaire. In each case, the figures in column 1 relate to seminar participants (P) and those in column 2 to respondents who did not attend the seminar (NP).

Table 5.1: Pre-seminar questionnaire: extent of previous use of OHP and reasons for non-use

	P	NP
	n = 8	n = 11
1a **Have you ever used an OHP in your teaching ?**		
a. No, never.	4	3
b. Yes, once or twice.	2	2
c. Yes, occasionally.	1	3
d. Yes, frequently.	1	3
If you answered (a) go on to Question 1b. If you answered (b), (c) or (d), go on to Question 2.		
1b **So you have never used an OHP. Why not ?**		
a. I've never felt the need.	1	3
b. I don't know how to use an OHP.	4	-
c. There isn't usually a machine available.	-	-
d. I don't know where the transparencies are kept.	1	-
e. I don't like using machines.	-	-
f. (other) ...	-	-

Questions 2 and 3 were answered only by those who had used the OHP before.

Table 5.2: Pre-seminar questionnaire: purpose of OHP use

2. **How many times have you used the OHP this term ?**	P n = 4	NP n = 8
a. Not at all.	2	6
b. Once.	1	-
c. More than once.	1	2
3. **For which of the following purposes do you (normally) use the OHP ?**		
a. writing new words and phrases during the lesson	-	2
b. writing students' errors during the lesson	-	-
c. exercises prepared before the lesson	1	5

d. answers prepared before the lesson	-	2
e. answers given during the lesson	-	3
f. reading texts photocopied from newspapers, etc	1	1
g. gapped texts (e.g. songs)	-	3
h. pictures: e.g. photocopied cartoons, pic. stories	1	5
i. maps	1	-
j. pictures drawn during the lesson	-	1
k. questions/instructions for an activity	-	2
l. games	-	2
m. crossword puzzles	-	3
n. (other)	-	-
- presentation of grammar rules	1	

Question 4 was again answered by all respondents.

Table 5.3: Pre-seminar questionnaire: attitude to OHP use

4. Tick the box(es) that describe your attitude to the OHP.	P n = 8	NP n = 11
a. Don't really see how it can help.	-	-
b. Haven't got time to prepare materials.	1	1
c. Might use it if I knew how to.	4	2
d. Have had bad experiences with OHPs in the past.	2	-
e. Useful substitute for board.	1	1
f. Useful supplement to board.	4	8

Responses to Questions 1 and 2 suggested that only one of the participants was a frequent OHP user: four had never used an OHP and of the remaining three only one had made use of the OHP in the early part of the term (*cf* three frequent users among non-attenders). Q.3 revealed that when the OHP was used by seminar participants it was for a limited range of purposes (*cf* the non-attenders). Answers to Q.1b and Q.4 suggested that the four subjects who had not used the OHP before were open to instruction (*cf* the three non-attenders who said, in answer to Q.1b, that they had never felt the need to use the OHP).

On the basis of these responses, four objectives were formulated for the seminar. It would:

1. encourage participants who had never used the OHP before to do so;
2. encourage participants who had used the OHP before to do so more frequently;
3. encourage the latter group to use the OHP for purposes other than those for which they had previously used it;

4. facilitate 1-3 above by modifying participants' attitudes towards the OHP (i.e. developing confidence, encouraging comparison with other classroom aids, such as the board).

These objectives are reflected in the questions asked during the post-seminar interviews and represent criteria against which the effectiveness of the seminar can be judged.

5.2.6.2 Post-seminar interviews

All participants were interviewed in the third week after the seminar to ascertain short-term effects. Questionnaires were used as the basis for the interviews. A further informal check was carried out on six participants the following term to establish whether these individuals had used the OHP since the previous interview.

It had been anticipated (see previous section) that changes would be seen under four heads: use *vs* non-use, frequency of use, purpose of use and attitude. The results are summarised under these headings.

Use vs non-use

In Table 5.4, below, the letters A-H refer to individual participants. Previous use (i.e. use prior to the seminar, to any degree) is indicated by a plus sign (+), non-use by a cross (x).

Table 5.4: Previous OHP use

A	B	C	D	E	F	G	H
x	x	x	x	+	+	+	+

Frequency of use

In Table 5.5, below, plus signs indicate reported use. Columns 1, 2 and 3 refer to Time 1 (the early weeks of the term, pre-seminar), Time 2 (first interview), and Time 3 (second interview). Participants G and H were not interviewed at T3 since there seemed little doubt that their pattern of use could be extrapolated.

Table 5.5: Frequency of OHP use

A	B	C	D	E	F	G	H
T1 2 3	T1 2 3	T1 2 3	T1 2 3	T1 2 3	T1 2 3	T1 2 3	T1 2 3
x + +	x (+) +	x + +	x x x	x x x	x x +	+ +	+ +

As far as objectives 1 and 2 are concerned (that the seminar would encourage first-time use and increase frequency of use), participants A to F are of particular interest. At T2, A and C, previous non-users, claimed to have exploited the OHP for very specific purposes and were able to describe what they had done (see response (n) in Table 5.6, below), and two days later B confided that she had used the OHP the previous evening and that it had worked well; at T3 all said that they had continued to make use of the aid. F, who had used the OHP previously, also seemed likely to continue to be at least an occasional user. D, who had not used the OHP previously and E, who had, despite professions of interest, had not acted on that interest.

The possibility exists, of course, that if OHP use has been frequent but indiscrimate, greater selectivity will lead to less frequent use. There is no evidence from the study that the seminar had had that effect (see Table 5.7, below).

Purpose of use

Column T2 in Table 5.6, below, indicates (self-reported) changes in OHP use following the seminar; Column TF, where F = Future, indicates projected uses.

Table 5.6: Purpose of use: aggregated results for the group of eight participants

	T1	T2	TF
a. writing new words and phrases during the lesson	-	-	1
b. writing students' errors during the lesson	-	-	1
c. exercises prepared before the lesson	1	1	4
d. answers prepared before the lesson	-	-	3
e. answers given during the lesson	-	-	1
f. reading texts photocopied from newspapers, etc	1	-	2
g. gapped texts (e.g. songs)	-	-	2
h. pictures: e.g. photocopied cartoons, pic. stories	1	2	5
i. maps	1	1	4
j. pictures drawn during the lesson	-	-	2
k. questions/instructions for an activity	-	1	-
l. games	-	-	1
m. crossword puzzles	-	-	-
n. (other)			
- presentation of grammar rules	-	1	1
- jumbled text	-	1	1
- text structuring	-	1	1
- grammar summary projected on to board; examples added on board	-	1	-

Reported uses prior to the seminar (T1) are clearly very limited, and it is not surprising that certain modest effects are visible at T2, nor that these effects are as modest as they are, given that the T2 interviews took place within three weeks of the seminar. TF suggests that the seminar may have had the desired effect in broadening awareness.

Attitude

Participants were also asked if their attitudes had changed as a result of the seminar and, if so, how. Seven of the eight said that their attitude had changed and Table 5.7, in which T1 is prior to the seminar and T2 after the seminar, indicates the nature of these (self-reported) changes. One person said that no attitude change had taken place. Since this person's attitude appears to have been positive prior to the seminar, no conclusions can be inferred from this response.

Table 5.7: Nature of attitude change

	T1	T2
increased confidence	n.a.	3
see OHP as substitute for board	1	2
see OHP as supplement to board	4	4
other:		
- *thought it was too complicated; now see it's worth using*	n.a.	1
- *insight into uses*	n.a.	2

5.2.7 Conclusions

It will be recalled that the evaluation had two purposes: to assess the effectiveness of the seminar and to consider what lessons might be drawn for future seminars in the programme.

Post-seminar evaluation offers some evidence that the objectives of the seminar had been at least partly fulfilled: i.e. that change in both behaviour and attitude had taken place. Three of the four subjects who had not previously used an OHP did so in the weeks following the seminar, and this was subsequently sustained. Four, including one who had previously had a bad experience with the aid, said that their confidence had increased; the range of purposes for which it was used had also broadened.

A week or so after the seminar, one of the participants commented that there had been a noticeable increase in OHP use among ML teachers. Question 10 in the post-seminar questionnaire sought to follow this up in an attempt to discover whether there had been any indirect effects of the seminar. Asked if they had talked about the seminar to staff who had not been present, four said they had. One had responded to a request for information; one had mentioned to a colleague that it had given her ideas; and a third had had an informal discussion about OHP use. This kind of ripple effect merits further exploration.

The final question in the post-seminar questionnaire (Appendix 5.2) asked how the seminar might have been improved. Several participants suggested that a hands-on component (an unrealised part of the original plan) and the provision of non-language-specific materials would have been appreciated. Support in these forms might have made a difference to D and E. One participant also commented that too much reliance had been laid in the early

part of the seminar on participants talking about their own experience when this was rather limited. Comments of this kind provide a useful basis for the planning of future seminars.

It is neither feasible nor, probably, desirable to evaluate all one's teaching in this way. However, one might argue that unless sampling exercises of this kind are carried out regularly, the relevance of that teaching and its value in terms of effects must remain open to doubt. Although the perspective taken here is essentially summative, there are lessons of a developmental nature to be drawn from the results of the study, as the previous paragraph has indicated. These are relevant not only to any future seminar on the same topic for a similar group but also to the teacher development programme within which the seminar took place and, indeed, to the planning of any practically-oriented INSET programme.

5.3 IMPLICATIONS OF THE CASE STUDY

The study did not try to ascertain changes in the degree of pedagogical awareness or technical skill demonstrated by the teachers as a result of the seminar. For this, classroom observation (pre- and post-seminar) would have been necessary.

Since its objectives were more limited, it was possible to use more limited (and less time-consuming) forms of data-collection. For instance, it relied for its post-measure on self-report elicited through interview, a procedure which, given the small numbers involved, took less than 1½ hours in total on the first occasion and no more than 30 minutes on the second. Self-report is, of course, a form of data-gathering which is potentially open to criticism on the grounds that respondents may wish to please the investigator or protect themselves from blame. In this case, there is good reason to suppose that the self-reports are reliable. I was not in a position of authority over the participants and had worked with several of them on a previous research project. The relationship was essentially that of professional colleagues. Moreover, nothing rested on the results. I began each interview by explaining the overall aims of the research, making it clear that the main focus of the research study was the evaluation instruments and not the success or otherwise of the seminar itself. Positive answers were, however, followed up, as indicated by response (n) in Table 5.5, above.

Evaluation instruments (written questionnaire, individual interview using standard questionnaire) were used in this study for two basic purposes:

1. to establish baseline behaviour and attitudes which would inform the content of the seminar (a descriptive, pedagogically-oriented purpose) and serve as a yardstick by which to measure subsequent change;
2. to assess the nature and extent of any changes in behaviour or attitude that might be directly attributable to the seminar, and to explain any non-effects (an evaluative, summatively-oriented purpose).

As the study has demonstrated, even such limited forms of evaluation as these can be revealing. Without the pre-seminar questionnaire or some other similar form of elicitation, the seminar would have been less sharply focused and evaluation of behavioural or attitudinal effects less reliable. Without the post-seminar interview at Time 2, there would have been no evidence of effects of the seminar (or, if the anticipated effects had not materialised, any indication of the reasons for this).

The pre-seminar questionnaire served its intended purposes. However, there is a limit to the number of questionnaires even cooperative colleagues can be asked to complete and for this reason the second questionnaire - based partly on the first - was administered orally. This produced the necessary comparative data while allowing for some flexibility in following up questions. It also ensured a complete data-set.

Special reference must be made to the informal interview at Time 3, which had not formed part of the original research-plan. Two factors influenced the decision to include a further set of interviews. In setting the time of the T2 interviews (in the third week after the seminar), it had been assumed that, provided subjects had had sufficient opportunity, they would have used the OHP; non-use during this two to three week period, on the other hand, would indicate that they would probably never use it. It became apparent from the interviews at Time 2, however, that this assumption was not necessarily well-founded. Two of the four subjects who had not used the OHP before had still not used it, but both said they planned to; the fact that intention was converted into action in one of these cases

has already been referred to. A similar positive attitude was shown by one of the subjects who had used the OHP before (F), but only occasionally. She had not used the OHP since the seminar, but again said that she planned to. It therefore seemed logical to extend the period of the study to determine whether good intentions on the part of non-users became anything more than that.

The second factor to play a part was the realisation, prompted by a rereading of Tyler (1949), that any positive effects observed at T2 might be short-term, the equivalent of playing with a new toy which is discarded when it has lost its novelty value. In other words, without a further check, there would be no proof that any effects observed at Time 2 were long-lasting. The study thus provides a practical illustration of the problem, noted in Chapter 2 (2.6.1), of the timing of post-programme effects.

CHAPTER 6

CASE STUDY 2

A REFRESHER PROGRAMME FOR ESOL TEACHERS

6.1 INTRODUCTION

The focus in Case Study 1 was on the collection and subsequent use of baseline data; this chapter looks at in-course and end-of-course evaluation in the context of a short open-access programme of the 'refresher' type.

To judge by the questionnaire responses reported in Chapter 4, a variety of measures are used for in-course and end-of-course evaluation on non-certificated open-access programmes (see Appendix 4.4). However, the interviews suggest that some institutions running short programmes of this type do relatively little to evaluate them beyond administering an end-of-course questionnaire and holding end-of-week review meetings (Informant 2: 20-22; Informant 3: 3; Informant 5: 9, 12). The interviewees may have been unwittingly understating the case; if they were not, reasons for the lightness of touch and the tendency to rely on end-of-course evaluation are nevertheless not hard to find. Refresher courses typically run under pressure, as informants testified: tutors want to squeeze as much into the course as possible, participants want to get as much out of the course and their time in the country. Both may resent the encroachment of evaluation on class time during a course, and participants may be reluctant to spend time on evaluation-related activities out of class. The end-of-course questionnaire, on the other hand, has something of the status of a convention. Participants expect to be asked what they thought of a course and tutors and managers need some reassurance that participants are satisfied. Although there is a recognition that end-of-course evaluation is suspect (Informant 2: 15-16 but *cf* Informant 5: 49), the perception that no more can be done because of time constraints (Informant 2: 16, 34; Informant 3: 4; Informant 5: 5) may inhibit both

consideration of other possibilities (which might prove more useful for programme development purposes) and review of what is done.

The case study presented in this chapter is an examination and assessment of the evaluation procedures used on a short programme which has been intensively evaluated since its inception over ten years ago. The research questions (see below) prompted the decision to carry out a longitudinal study of the programme rather than a cross-sectional analysis of procedures in place at the present time.

Research questions

1. *What evaluation procedures* ***other than those normally used*** *for a programme of this kind have proved feasible and useful ?*

2. *Are different (or fewer) evaluation procedures needed as a programme becomes more established ?*

6.2 THE STUDY

6.2.1 Introduction: TLE and its origins

IALS is a self-financing institution and in its early years there was a pressing need to establish a profile in the ELT marketplace and a financial base. Attempts were made to attract students during the academic year, from October to June, but the main effort went into the design and promotion of summer courses, since it was assumed (rightly) that initially at least these would be the principal source of income. Most of these courses - General English, Spoken English, English for University Studies - were for language learners, but teachers were also seen as a potential market.

The first venture in this direction was a three-week summer course for NNS EFL teachers, somewhat unimaginatively named 'Teachers of English'. This was run twice in 1980 and twice in 1981. These courses, designed and directed by an outside Course Director, were not wholly successful. As the Course Director wrote at the time:

> Taking all the comments in this report into account, the Institute may conclude that certain changes or alternatives should be examined in terms of (1) the general perspective, especially on language teaching, that the courses have adopted, and (2) the balance of components, to cater for those teachers more inclined towards proficiency or towards language teaching as their main reason for attending the course.
>
> (Higgens 1981)

Early in 1982 I was invited to redesign the methodology component of the programme and the new version, subsequently to be named 'Teaching and Learning English' (TLE), was planned in conjunction with B. Heasley and launched in summer 1982. Since then, the programme has run continuously, normally three times each summer. In the early years, there was a maximum intake of 24 per course. After a period during which this was reduced to 20, it has now been allowed to 'float'.

The scope and objectives of the programme, as described in a recent brochure, are as follows:

> TLE is a wide-ranging refresher course for teachers who wish to extend their knowledge of modern developments and techniques, familiarise themselves with recent ELT materials, and improve their fluency...
>
> **Course objectives**
>
> The course aims to illustrate current practice in ELT and develop participants' confidence in using English for teaching and other purposes.
>
> (Summer brochure 1995)

Morning sessions provide language practice (through activities also designed to illustrate specific pedagogic techniques); raise awareness of aspects of the language and culture that are relevant to participants as learners as well as teachers; and deal through tutor presentations and directed discussion with a wide range of methodological topics. This has been the basic formula since 1982. The nature of afternoon sessions has been more variable, particularly in relation to the balance between tutor-led sessions (e.g. talks, workshops, video-viewing) and self-directed work (e.g. group preparation of lesson-plans or individual work on self-access resources and tasks related to these). Currently, afternoon sessions include workshops, with topics selected by participants from a short menu; videos about the language and/or language teaching; and opportunities for self-

access work. In its mix of language improvement and methodology, TLE is probably typical of the short refresher programmes run by U.K. institutions. One of the ways in which it may differ is in its commitment to the needs of the individual, manifested not simply in the provision of resources, time for self-access study and individual tutorials, but also in an emphasis on self-evaluation. A second way in which it may differ is in the extent to which it has been evaluated.

6.2.2 Overview of the case study

Given the problems diagnosed by the previous Course Director, it was logical that the new programme should be carefully evaluated, in its early stages at least, and the proposal for TLE 1982 (McGrath 1982) included detailed specifications for the instruments and procedures to be employed for the purposes of programme evaluation (see Appendix 6.1). Some of these were later modified, and new measures introduced (see 6.2.3, below). Although other measures and criteria were also used to evaluate the success of the programme (see Appendix 6.2), the primary objective of these multiple measures was the assessment of participant satisfaction. The data gathered served both formative and summative purposes.

Section 6.2.3 draws on available documents (the course proposal already referred to; publicity literature; course reports for the years 1982-92; standard pro-formas; miscellaneous items such as copies of participants' study diaries and post-course letters; academic papers on aspects of the course) and the experience of myself and other former CDs to describe the evaluation measures and the way in which they were used; and section 6.2.4 evaluates these measures. Section 6.3 then reports on an experiment in 1993 into the effect of discontinuing one of the measures. The conclusion argues for the need to take periodic stock of evaluation procedures and to ensure that course documentation is full and complete.

6.2.3 Evaluation procedures: 1982-1992

The initial battery of **in-course** evaluation procedures consisted of individual tutorials towards the beginning and end of the course, rapid assessment forms (RAFs), trouble-shooting sessions in two forms, and tutor meetings (see Appendix 6.1). Three new measures were subsequently introduced: a study diary, a weekly review and an in-course

letter. **End-of-course** evaluation consisted of participant self-evaluation (oral) based on individual statements of objectives formulated on the first day of the course, an end-of-course questionnaire and whole-group discussion, and a tutors' meeting (see Appendix 6.1). A report containing recommendations was then written by the Course Director (CD) and submitted to the Director of Studies. Although the form of the questionnaire was modified to answer different questions at different stages of the programme's development, the basic end-of-course procedures remained stable throughout the period 1982-92, with the occasional addition of written tutor reports as input to the CD's report; in-course procedures, as will become clear from the discussion below, were more variable. A description of the rationale for each of these procedures and their outcomes follows.

6.2.3.1 Individual tutorials

As indicated in Appendix 6.1, individual tutorials were scheduled twice during the course. Initially, the first set of tutorials were intended to guide and subsequently stimulate reflection on the ways in which course activities and course processes could be adapted to participants' own teaching circumstances. Their timing, during self-study sessions, was deliberate. The second set of tutorials took place in the final week of the course. Although ostensibly serving the needs of participants (an opportunity to get advice on how they could follow up on their return to their own countries), these tutorials were also seen as being an opportunity for tutors to assess the impact of the course on individuals.

The approach to the tutorials soon changed. Individual tutorials gave way to small-group tutorials during self-access periods, the basis for the groups being similarity of teaching context. These prepared for unsupervised cooperative work on lesson-planning using ideas and materials suggested during the course (Hamp-Lyons 1986). Later, there was a return to individual tutorials. These were timetabled for the second afternoon of the course and used for two new purposes: to discuss the realism of participants' individual learning objectives, as formulated on the first day of the course, and to elicit their attitudes to the course thus far. In this latter format, which still survives, the tutorial was envisaged by tutors as serving three purposes: it was (a) a way of evaluating the match between individual and programme - a kind of early warning system (b) an opportunity to deal with any misunderstandings if these became apparent (effectively a matter of changing the individual to suit the programme) and (c) a means by which individual needs or interests which fell outside the scope of the programme could be ascertained and, if possible,

furthered (a purpose that can also be fulfilled by diaries - see Appendix 4.6, Informant 1: 22). In due course the second set of tutorials were made optional on the afternoon of the penultimate day, in keeping with a more explicit focus on self-direction.

No records survive of tutorials and it is therefore difficult to assess how far they followed the intended format and served the intended purposes. Oral recollection by former CDs suggests that in latter years purposes (a) - (c) have all been served. The strongest evidence for this relates to the first purpose and takes the form of transfers to other courses where the mismatch between participant expectations and programme has been irreconcilable (a process that might have taken longer otherwise).

6.2.3.2 Rapid Assessment Forms

The RAF used on TLE was devised by the present writer in 1982 and was used on all courses throughout the period of the main study (1982-92) to provide detailed feedback on individual sessions. Table 6.1, below, contains an extract from such a form.

Table 6.1: Extract from Rapid Assessment Form (1986)

EVALUATION OF WEEK 2
Try to fill in this form daily and hand in the completed form at lunchtime on Fridays. Please use the code shown on the right, adding an explanation if you gave a score of 2 or lower.

INTEREST (I) / VALUE (V)
5: very interesting/very useful
4: interesting/useful
3: quite interesting/quite useful
2: not very interesting/useful
1: boring/useless

	SESSION 1 9.00-10.00 **Language Activity**		SESSION 2 10.00-11.00 **Language Teaching**		SESSION 3 11.30-12.45 **Language Focus**		SESSION 4 14.00-15.30 **Workshop or Video**		**Additional comments**
	I	V	I	V	I	V	I	V	
Monday	Listening Activities		Teaching Listening		Text Study 1 Listening				
	5	5	5	5	5	5	5	5	
	4	4	4	4	4	4	4	4	
	3	3	3	3	3	3	3	3	
	2	2	2	2	2	2	2	2	
	1	1	1	1	1	1	1	1	

Participants were asked to complete the form daily, rating each session for interest (I) and value (V) separately using a scale of 1 (= boring/useless) to 5 (= very interesting/very valuable) and commenting on any rating below 3. In a later modification to this procedure they were also asked to add tutors' initials for sessions where the same material was being taught in split groups.

The RAFs were handed out, and their purpose explained, as part of the course introduction on the first morning of the course. Participants were reminded to fill them in on at least two occasions thereafter (typically at the end of the first day and at the beginning of the following day). Anonymity was permitted, but many participants added their names. The forms were collected at the end of classes on the Friday morning of each week, tallied, and discussed at the weekly tutors' meeting on the Friday afternoon.

Despite the cautionary remarks of Saville and Andrews (quoted in 3.2.5), the RAFs proved to be useful in a number of ways.

Initially their primary contribution was to programme development. When participants' scores had been tallied and an average calculated (the tally divided by the number of participants) for each session they provided a rough and ready indication of session *interest* (assumed to relate to either to the intrinsic interest of the topic or the tutor's treatment) and *value* (construed as perceived relevance of content). Sessions scoring an average of 3.5 or higher (out of a possible 5.0) for interest and value were judged to be satisfactory. Sessions attracting averaged scores lower than this were discussed at the weekly tutors' meetings and participants' comments on these sessions and tutors' self-evaluations taken into account in making draft recommendations. Where changes to the timetable or materials were strongly indicated, action was taken on these for the following course. In most cases, however, decisions requiring radical changes were taken only at the end of each series of courses (i.e. at the end of each summer), when results could be compared.

Table 6.2, below, shows the averaged ratings for the same two methodology sessions across three courses.

Table 6.2: Extract from summary of RAF ratings for Methodology sessions: comparisons across courses (1983)

SESSION	COURSE		5	4	3	2	1	AVERAGE OF RESPONSES	TUTOR
Language Teaching	1	I	4	4	5	1	0	[14] 3.7	IM/LG
		V	1	3	7	2	0	[13] 3.2	
	2	I	2	7	5	5	0	[20] 3.2	IM/DW
		V	3	4	6	6	1	[20] 3.1	
	3	I	2	2	10	0	0	[14] 3.4	IM/LG
		V	4	0	9	1	0	[14] 3.5	
Oral Practice: Visual Stimuli	1	I	9	7	0	0	0	[16] 4.5	IM
		V	12	2	2	0	0	[16] 4.6	
	2	I	8	7	4	1	0	[20] 4.1	IM
		V	5	11	3	1	0	[20] 4.0	
	3	I	7	4	3	1	0	[15] 4.1	IM
		V	8	3	3	1	0	[15] 4.2	

Using this kind of focussed summary and particularly the figures in the column *Average of Responses* (where the figure in square brackets represents the number of participants) permits comparison of session appeal across courses within the same year or across years, thus facilitating decisions concerning the necessity for revision of the content of the course or the procedures adopted. With hindsight, it is apparent that if a column were also added for notes on changes made as a result of these decisions (e.g. OHP transparency added; activity 2 replaced), it might be possible to assess the effect of these changes.) However, as the entries for Oral Practice (Visual Stimuli) bear out, there will always be a degree of variation, even when the materials and tutor are held constant. This argues for the exercise of some caution in making changes, at least when a reasonably successful formula seems to have been achieved.

A second way in which the RAFs were used was as a device to monitor individual levels of satisfaction during the course. For this purpose, the explanatory comments that accompanied any negative ratings were a helpful adjunct to the ratings, indicating that on occasion the source of the problem was located within participants themselves ("I felt tired this morning. Wasn't able to concentrate."). Where appropriate, and if the individual could

be identified, negative comments and patterns of ratings below 3.0 were followed up on a one-to-one basis; other comments were picked up by the tutor leading the first session at the beginning of the following week.

Mention should also be made of the inclusion of tutors' initials. This was intended to provide feedback on the relative success of different teaching styles and approaches to the (standard) material - a course development objective, but also potentially a means by which tutor performance (or popularity) could be evaluated. In the early years of the programme, when tutors tended to be permanent members of staff, this feature was seen as an instrument for self-evaluation; it was dropped from the RAF returns once the course development objective was felt to have been achieved and staffing patterns changed.

As TLE became more established new sessions were introduced in response to emerging issues within the field and materials and procedures were modified as a result of tutor initiatives, as the following quotations indicate:

> The course content was similar to that in previous years.... In Blocks 2 and 3 alternative approaches were tried to a number of sessions and notes on these are held in the session files. In Block 3, SD and JH introduced an addiitonal session on 'Teaching Mixed Ability Classes' at the request of Ps and material for this is now on file.
>
> (Course report 1987)

> The content was changed from that of 1988 (either in terms of new material for 'old' topics or as new topics) in the following sessions [14 sessions listed]
>
> (Course report 1989)

In relation to the new sessions, the RAFs served their original purposes (formative and summative); in relation to existing elements, the emphasis was on ensuring that the standards of satisfaction that had been achieved previously were maintained. At the level of component sessions, RAF scores were ideal for this purpose.

6.2.3.3 Tutor meetings

Tutor meetings took place on a weekly basis, on the Friday afternoons following teaching in the first two weeks of the course and on the final day of the course. The purpose of the first two meetings was to review the week and prepare for the next; in the final meeting the course as a whole was reviewed. Each meeting followed a similar format: tutors discussed their feelings about the course and raised any specific points concerning timetabling, materials, resources, individual participants or the group as a whole; participant feedback from the RAFs and other sources was also discussed. Notes were kept of the points made and these formed the basis for action the following week or recommendations for future courses. The following extract from a course report indicates the relationship between tutor meetings, action and input to future planning:

> **Language Learning**: Ts [tutors] felt discussion sheet was ambiguous. It was revised for Blocks 2 and 3 and worked better then.
>
> **Spoken English 1**: Both Ps [participants] and Ts were unhappy with this session. Ts felt it worth having TLE compilation audio and video tapes for all the material used. Ps felt pressed for time, inadequate in identifying error in each other, sensitive about peer criticism on Day 2, the need of a model, the need for security of NS criticism of performance.
> **Materials Evaluation**: Ps find this a difficult task. In Blocks [i.e. courses] 2 and 3 we reduced the textbooks surveyed to 2 titles, but Ps still found the task difficult.
> (Course Report 1987)

6.2.3.4 Study diary

During the course introduction on the first day of the course, the point was made that participants could learn not only from the content of the course but also from experiencing the course in the role of learners. To this end, they were recommended to keep a study diary in which they recorded their attitudes to what happened during the various sessions and the insights gained into learning and teaching. They were taken through examples of diary entries and given a pro-forma allowing space for five entries on a double-sided A4 sheet. One desired effect was that learning would result from a process of honest self-evaluation (see also 6.2.3.4, below). Although there was no compulsion to keep the diaries, deliberate reference was made to them during the first week and a supply of Study Diary pro-formas was pinned to the wall of the teaching room. In the final week of the

course it was suggested that participants who had kept diaries might wish to hand them in as input to individual tutorials.

Appendix 6.3 contains an extract from a study diary. It will be clear from this that although the primary focus is on personal gains, as intended, such a diary also contains incidental comment, implicit or explicit, of an evaluative nature. Thus, we learn from this diarist that:

a. most of the games were novel and some were considered usable;
b. the language analysis work had raised the writer's awareness in ways that had implications for the classroom;
c. the work on vocabulary had prompted him to consider devoting more time to this in his own teaching.

The entries also suggest the following:

d. participants need advice on how games can be incorporated into a language course;
e. more time could usefully be allocated to discussion/demonstration of the use of video.

Potentially, this is information which can be compared with that available from other sources or obtained by other means. In practice, this kind of comparison probably did not take place. The CD report for 1984 included a quotation from a diary with the comment 'all are encouraged to keep such a diary, few do'; oral recollection by former CDs suggests that this continued to be the case after 1984. The emphasis in the presentation of the diary on its being for the benefit of the participant and its optionality were no doubt factors in this.

6.2.3.5 First-week letter

Between 1983 and 1986 participants were set a week-end task at the end of the first week. This was to write a letter to the Course Director explaining that they had been suddenly called back to their own country, and commenting briefly on what they felt they had gained from the week. The letters were collected on the Monday morning. From a language point of view, this was writing practice; from a pedagogic perspective, it was intended to prompt review and reflection; as regards evaluation, it was an attempt to complement the analytical quantitative data from the weekly RAF sheets with a more

global view which would allow an outlet for individual impressions and preoccupations. The Course Director responded to some of the points arising in a plenary session the following morning; others were dealt with individually.

The CD report for 1983 notes that the first-week letters had 'revealed that several participants had not expected to be in such a large group, and felt reluctant to speak out during plenary discussion (the problem being one of shyness rather than language ability *per se*)'. There is no indication whether any immediate action was taken on the basis of this insight. However, the letters were clearly an early warning that group size and session management was an issue, a point which also surfaced in the end-of-course questionnaires and whole-group discussion, when participants specifically 'expressed a wish for more small group discussion' and 'suggestions included: "permanent small groups" and "don't take so many people"'. The report goes on to discuss the relationship between group size and responsiveness and puts forward some of the options for dealing with the problem.

The letter-writing task was inadvertently not set in 1987 and, despite CD reports in 1987 and 1988 drawing attention to its value (see 6.2.4.3), it thereafter disappeared.

6.2.3.6 Trouble-shooting sessions in Weeks 2 and 3

It was anticipated from the outset of the programme that opportunities would be needed for tutors to deal with any tensions that arose on an interpersonal level between individuals or subgroups or in relation to the course itself and for participants to raise teaching-related issues that would not be dealt with as part of the standard course content. The Integrated Activity, a role play in which participants in small groups took turns to be tutors or course participants (see Appendix 6.4), was intended to surface and defuse any problems; and what was originally called a Forum session, in which participants' prepared questions on pedagogic issues formed a springboard for general discussion, was planned to minimise any dissatisfaction with course coverage. The Integrated Activity was a method of involving the participant group, in a pedagogically appropriate manner, in the solution of emerging problems, the Forum an attempt to anticipate problems. Both were therefore intended to serve as in-course monitoring devices. The Integrated Activity took place in Week 2 and the Forum in Week 3.

The Integrated Activity receives no mention in CD reports and since course timetables are not available for the earlier courses it is unclear when or why the activity was dropped. Its lack of any visible value as an evaluation instrument may have been one factor in its being discontinued, another may have been its marginal relevance - as an example of a type of oral fluency activity - to classroom teachers.

CD reports contain only one reference to the Forum. This follows a table showing that the Forum was felt by six (of the 24) participants to have been 'of very little value' and reads: 'a general discussion of participants' teaching problems was replaced [in Blocks 2 and 3] by a demonstration of the Lothian Region's Graded Objectives Scheme' (Course Report 1982). Despite its swift demise in 1982, the Forum made occasional reappearances (CD recollections), but apart from one surviving set of participant questions (see Appendix 6.5)) no records were kept of either inputs or outcomes.

6.2.3.7 Participant self-evaluation

Unlike the other measures described, participant self-evaluation bore a more indirect relation to programme evaluation. Managed through a complex of interlocking devices, it evolved over the first two-three years of the programme. On the first morning of the course, participants completed a pro-forma which guided them towards a prioritised formulation of language learning objectives (and the means through which these could be achieved) for the duration of the course (Appendix 6.6 contains extracts from this pro-forma). The objectives were then discussed in an individual tutorial the following day, pursued (or not) by the individual, progress recorded (or not) in the Study Diary, and achievements self-evaluated on the final day of the course in response to oral or written prompts (now part of the end-of-course questionnaire - see Appendix 6.7). The premise underlying this approach was that if participants were encouraged to formulate realistic objectives for themselves a sense of achievement (and therefore satisfaction) in relation to the language component of the experience was a more likely outcome than otherwise; there was also a recognition that it is important to be aware of participants' expectations.

CD reports between 1988 and 1992 consistently refer to this aspect of the evaluation findings, sometimes in tabular form (as in Appendix 6.2) and sometimes in selective summary form, as in the example below:

> 72% felt that they had achieved their language learning objectives during the course. Most of the rest suggested that they failed to do so because their objectives were unrealistic or they did not devote enough time to them. A pleasing 90% felt that they would be able to cope better with classroom problems they had identified.
>
> (Course Report 1988)

Prepared for in the way described above, end-of-course self-evaluation can preview and complement participant evaluation of programmes, making it more likely that the cause of any shortcomings in the experience is correctly identified. As Hamp-Lyons (1986) put it: 'participants understand themselves as learners a little better as a result of this self-analysis, and can see what unfulfilled objectives are their own responsibility and which the course's'.

6.2.3.8 End-of-course questionnaire

As far as content is concerned, the end-of-course questionnaire currently used (see Appendix 6.7 for 1992 version) probably differs little from that used in other institutions. Although no records survive of the versions used on the earliest courses, the course reports suggest that, apart from a question asking whether the course represented value for money, the bulk of the questionnaire was given over, as it is now, to questions on the balance between theory and practice, the value of the major components, the sessions felt to be most and least useful, and the adequacy and use made of resources. All of these questions yield quantitative data that can inform decision-making and the resulting figures are consistently referred to in course reports as evidence of the programme's success in meeting participants' needs.

The following is typical:

> ... the course appeared to be well-received across the three blocks. The balance of theory and practice was perceived by the vast majority to be right although there was a slightly higher number in Block One who felt there was too much theory. ... everyone considered that they had received valuable new ideas from the language activities and no individual session was felt to be pedagogically valueless by more than a few participants.
>
> (Course Report 1992)

What is perhaps a little unusual about the questionnaire is the form in which it is administered and the way in which results are used. The first part of the form, which is broadly concerned with level of participant satisfaction, is handed out on the penultimate day of the course with the instruction that it is to be completed for the following day. The completed forms are collected when participants arrive the next day and the results collated and transferred to an OHP transparency by the CD. After a coffee break, time is allocated for small-group discussion of the second (more general part) of the questionnaire, which relates to expectations and whether participants' objectives have been achieved, and this is then completed individually in writing and collected. The results of the first part of the questionnaire are then presented to the plenary group by the CD and comments invited (see ***Whole-group discussion***, below). The potential advantages of this approach to end-of-course evaluation is discussed in 6.2.4.2, below.

6.2.3.9 Whole-group discussion

In its most consistent form, whole-group discussion for purposes of programme evaluation has been associated on TLE with end-of-course evaluation. The format has been unchanged since 1982.

On the final day of the course, following the collation of the quantitative results of section 1 of the end-of-course questionnaire, the CD presents the results on the OHP and leads discussion on any matters arising. In relation to participants, the intention of this procedure is twofold: it enables individual participants to see how far their own views are shared by others in the group and it offers an opportunity for those expressing minority views to explain these. From an evaluation perspective, whole-group discussion of the results is essentially a procedure which serves to corroborate and supplement the results of the end-of-course questionnaire: it allows for clarification of any unusual or puzzling questionnaire responses and checking that the obvious interpretation is also the correct one; and it allows participants to express their views on any other aspects of the course experience which they feel to be relevant.

On the occasions when ***weekly reviews*** were held (and no records exist of when these took place or what they produced) they fulfilled a dual purpose. Tutors led the discussions by indicating the interrelationships between sessions and then elicited participant reactions. Although the oral recollections of former CDs suggest that the reviews did produce occasional useful insights there is no evidence in the course reports that they influenced subsequent decision-making.

6.2.3.10 Course Director's report

The CD report is potentially *the* key element in institutional programme evaluation and review. The report synthesises evaluation data collected from various sources, by various means and at various stages, relates this to the circumstances under which the course took place, and makes judgements concerning their significance. If the programme is well established and full course reports are available for previous years, the resulting recommendations will be supported by factual data indicating the extent to which participants differed (in prior training, teaching experience, language proficiency, age, gender, teaching situation, expectations, personality, etc) from previous groups. The

evidence supplied by the report is the basis for action, and the cumulative evidence of the reports is an important input to programme review.

Comparison of the TLE reports for the period 1982-92 reveals the following:

1. formats vary considerably, rendering cross-course comparison more difficult;
2. details of participant data range from a simple statement of numbers (see 3, below and Appendix 6.2) to a course-by-course breakdown of nationalities, gender and teaching situation and discussion of the ways in which these differed from previous years and the effect of these differences;
3. timetables are included in only a small number of cases (a note in the 1984 report indicates that the timetable was stored separately with the course records (i.e. data on participants, full evaluation data and course materials) - unfortunately, these course records have not survived;
4. evaluation data varies from selective reporting in the early years to the reproduction in later years of full quantitative results for the end-of-course questionnaire and comments on these;
5. recommendations cover a very wide range of topics (e.g. admissions policy, group size, staffing policy, pricing, promotional materials, course intensity, methodological approach, course content, time allocation for CD administration, time allocation for course development, equipment); in some cases, recommendations narrowly concerned with materials development have clearly been filed separately;
6. innovations and changes are frequently described but these are not systematically related to recommendations in previous reports.

Responsibility for decisions on the recommendations of the report rests with the Director of Studies, to whom it is first submitted, the Director of the Institute, or the subsequent CD. In the absence of 6, above, it is unclear how CD reports have influenced decision-making and where they clearly have not, to what factors this can be attributed.

6.2.4 Evaluating the evaluation procedures

An attempt was made in the previous section to indicate both the actual and potential value for evaluation of the instruments used during the period covered by the study. In

some cases, even the potential was small (e.g. the Forum) or incalculable (the Integrated Activity); in others, the gap between potential and actual was large because the instrument was not used with any degree of consistency (e.g. the first-week letter). However, the principal difficulty in assessing actual value lies in the incompleteness of the available documentation. This is a point to which we return under ***Utility***, below, and in 6.4.3.

Such information as is available nevertheless permits certain conclusions to be drawn. Following the discussion in Chapters 2 and 3, these will be organised in relation to logistical factors, methodological factors, and utility.

6.2.4.1 Logistical factors

To judge from the survey, time appears to be the one reason why less evaluation takes place on short courses than on longer ones. The study suggests that this need not be a constraint if three conditions are met:

(a) the demands on participants strike a balance between in-class and out-of-class evaluation - in the case of TLE, RAF-completion, and the writing of first-week letters and study diaries all took place outside class time;

(b) in-class (and, if possible, out-of-class) evaluation is perceived by participants as being at least in part for their benefit - in-class examples being individual tutorials, self-evaluation, the Forum, the Integrated Activity; and out-of-class activities being the study diary and first-week letters);

(c) the demands on tutors' time are realistic - in this case, the burden of processing the RAFs and first-week letters fell to the CD, who had a time-allowance for administration; the results of the first part of the end-of-course questionnaire were tallied by the CD while another tutor managed the session; and tutors' meetings took place on non-teaching afternoons.

What posed the most obvious logistical difficulties were the individual tutorials. Various approaches to these were tried (e.g. rotation of sub-groups, so that while one group was having tutorials, another group was watching a video and a third was having a tour of the Institute's resources); all involved a tension between participant waiting-time, length of tutorial and staffing allocations. The course report for 1982 commented: 'the provision of individual tutorials, an innovation, appears to have been very much appreciated'. This is

not a reference to their function as an evaluation instrument, of course; but tutors' perceptions of their utility in this and other respects (CDs' personal recollections) have led to their survival in spite of logistical disadvantages.

6.2.4.2 Methodological factors

In relation to the evaluation of a single programme for institution-internal purposes, the most important methodological factors are reliability, validity, and comparability of results.

The potential *pros* and *cons* in methodological terms of certain of the instruments used on TLE (questionnaire, interview, participant diaries) were discussed and summarised in Chapter 3. One of the problems noted was the tension between reliability and validity with regard to the degree to which data-elicitation is structured. This is a particular concern when conclusions are based on individual responses; it becomes less of a concern when triangulation of source and method is possible, as indicated below.

Several of the instruments used to monitor participant satisfaction on TLE can be exploited in two ways. On the one hand, they yield data on an individual level (individual tutorials, first-week letter, RAFs) which can inform action at that level but from which there is no attempt to extrapolate. On the other, they permit data to be aggregated, informing decisions which either have implications for the group as a whole or the planning of the next course (first-week letters, RAFs). In other words, although all these instruments have certain obvious disadvantages in terms of reliability and/or validity at the level of the individual, when the results for a single instrument (such as the RAF) are aggregated, for a course and especially across courses, these results have some reliability; and when we look at the results obtained by using several methods, we have a multi-faceted picture of a course that has a kind of collective or overall reliability and validity.

The end-of-course evaluation procedures are designed to work together and are therefore also most conveniently treated together. The view has been expressed (see 6.2.3.7) that participant self-evaluation can help to make participant contributions to the process of programme evaluation more objective and therefore more reliable (Bailey and Ochsner 1983; Palmer, C. 1993). The procedures adopted for end-of-course evaluation (and

described in 6.2.3.8 and 6.2.3.9) also ensure that end-of-course evaluation is more reliable and valid than it might have been otherwise. Its relative reliability stems from the fact that participants' responses are considered: they have time to reflect on their responses to the first part of the questionnaire, and they have an opportunity to hear the views of others on the questions in the second part; the opportunity given to participants to comment on the results lends validity to their interpretation; and the open-ended invitation of 'any other comments' (in the questionnaire and during the discussion) permits other ideas to be aired. Of course, the possibility cannot be excluded that time, negativity, apathy or shyness militate against the expression of views both on topics covered by the questionnaire and those not explicitly mentioned.

As far as data analysis is concerned, the advantages and disadvantages of the various instruments are predictable from their formats. The results of the questionnaire, for example, are by and large easily quantified and therefore directly comparable with those from previous courses. This is also the case for the RAFs. The open-endedness of other instruments makes quantification and comparison more difficult. Given the dynamic nature of courses and the extent to which an experience is coloured by personal interactions, comparison across courses based on these instruments may even be impossible.

6.2.4.3 Utility

For various reasons, it is difficult to assess the *actual* utility of some of the measures employed.

Because evidence concerning the utilisation as evaluation instruments of the Study Diary, Forum, Integrated Activity and First-week Letter is minimal it is inconclusive as far as their utility is concerned.

As regards the Study Diary, this is not surprising. The diary was not compulsory and since it was not presented as an evaluative device there was no particular reason why participants should have handed it in. In the Forum, evaluation was very much part of the hidden agenda and it is unfortunate, with hindsight, that these sessions were treated as one-off events. If records had been kept of the questions and patterns been discernible in the

issues raised, this would have been a useful input to course review. Evaluation was more to the fore in the Integrated Activity. Indeed, problem scenarios were on occasion specially written to suit individuals on a course (CD recollection). Any attempt to assess the value of the activity would have been extremely difficult, however, since it would have required records to be kept of the problems selected by individuals, the discussions in which they were involved, and any subsequent effects, behavioural or attitudinal.

Like the Study Diary, Forum and Integrated Activity, the first-week letter, as noted above, was intended to serve a pedagogic purpose - in this case, to encourage participants to review materials and notes for the first week (i.e. a purpose equivalent to a class-based weekly review). Its general evaluative purpose was to ascertain the attitude of individuals at the end of the first week to the course as a whole (*cf* the RAFs, which provided feedback only on component sessions); at a more specific level, it was intended to reveal what was dominant in participants' minds. Although no letters have been preserved, it is evident from the extract from a course report reproduced in 6.2.3.5 that the letters did serve this specific evaluative purpose; and as a later report indicates, this was not an isolated case.

> Unfortunately this year we omitted asking Ps [= participants] to undertake the letter-writing exercise at the end of Week 1 and the weekly reviews. These omissions, I believe, were a loss to the course as we missed a further opportunity of encouraging Ps to consider how different parts of the course inter-related with each other, and of surfacing "issues" that were developing, such as unhappiness with the presentation groups...
>
> (Course report 1987)

Ironically, the implicit recommendation seems not to have been heeded, and we find a similar comment in the report written by the same CD for the following year: 'letter and weekly review - reinstate these and reviews in Weeks 2 and 3'. They were not reinstated.

The argument for individual tutorials, as used on TLE, as an evaluation instrument relates to their counselling function. Attempting to increase their evaluative utility by making tutorials in the final week compulsory and using these tutorials more explicitly to elicit participant feedback would have obvious logistical drawbacks and might simply produce data very similar to that generated by existing final-day measures.

The utility of other measures is perhaps more obvious. The RAF, despite its drawbacks as a precise measuring instrument, proved a useful means of tracking general levels of participant satisfaction with individual sessions within and across courses. The end-of-course questionnaire, combined with whole-group discussion and self-evaluation, provided a broader view of levels of participant satisfaction; tutor meetings were able to place and interpret these findings in relation to tutors' experience of the course and the programme.

When findings are positive, the only problem is one of complacency; when they suggest that change is necessary, then action has to be sanctioned. Reservations have already been expressed concerning the CD reports. The information required for comparative purposes is frequently not available (in the report) and there is seldom any clear indication of the action taken on recommendations (exceptions would be the report for 1987, which notes that 'extensive new tutor notes were prepared for this year and sessions reordered and materials revised according to recommendations from CDs in 1986', and that for 1990, which states that 'revisions were made to the material, most following the recommendations of 1989, viz ...' and goes on to list 12 sessions to which changes were made). An outsider would probably conclude that the reports were used primarily for purposes of accountability and that their potential contribution to programme development was realised only erratically. CDs' oral recollections suggest that much development work prompted by evaluation (content replacement; material/activity replacement and modification; strengthening of theoretical component through provision of hand-outs and self-access materials) has simply not been recorded in the reports.

6.2.4.4 Conclusions

The study points to three conclusions:

1. Where participant numbers are relatively small (e.g. up to 20) and there is a tutor to participant ratio of, say, 1 : 8, programme evaluation can go beyond the basic end-of-course questionnaire and weekly review without undue demands being made on participants or tutors.
2. Apart from end-of-course questionnaires, reviews and such other fairly commonly used procedures as individual tutorials and tutor meetings, consideration should also be given on the basis of their utility to the use of RAFs, first-week letters and self-

evaluation. Study diaries, if presented to participants in such a way that they generated evaluation data and access to this data could be assured, might also prove useful, as indicated in Appendix 4.6 by Informants 1 and 4 (but see also Chapter 7). Each of these four instruments could contribute to programme development; RAFs and self-evaluation (as described here) can also be used in summative evaluation.

3. Full documentation is needed of all forms of evaluation (pro-forma if appropriate, procedures, results, and how the results were used); these, together with a record of action taken on recommendations, are essential for purposes of programme review. Moreover, for management and programme personnel, it is important that summative information is seen to be used for formative purposes.

Incomplete though the evidence is, the study thus provides under point 2 above a suggestive answer to the first of the research questions ('What evaluation procedures other than those normally used for a programme of this kind have proved feasible and useful ?'). In the next section, we consider the second research question: 'Are different (or fewer) evaluation procedures needed as a programme becomes more established ?'.

6.3 REVIEWING EVALUATION PROCEDURES

6.3.1 Introduction

A rational explanation for the changes that took place in the period 1982-92 to the evaluation of TLE would be that these were the result of review. While this may be true of the first year or two of the programme, when the introduction of the first-week letter and modifications to the timing and purpose of the tutorials were clearly attempts to refine the battery of procedures, there is no evidence of any subsequent review of evaluation procedures.

Writing about routine in teaching, Maingay (1988) makes the point that

> ritual teaching behaviour ... is unthinking.... it is teaching behaviour that has set into patterns.... This kind of teaching is ritual in the sense that although there may be principles behind it, the teacher has never known, or has lost sight of, these principles, and is consequently going through the motions.
>
> (Maingay 1988: 118-119)

Evaluation can become equally unthinking and ritualised.

By 1992 TLE was firmly established. It ran three times each summer and consistently attracted a viable number of participants. Interestingly, it had survived in a recognisable form: the basic structure of the mornings remained unaltered, as did some of the materials; moreover, many of the evaluation procedures were still in place. It was high time to reconsider the necessity for such a battery of evaluation procedures.

6.3.2 The experiment: summer 1993

6.3.2.1 Background to the experiment

In planning for summer 1993, following discussion between those most directly concerned (including myself as Development Coordinator for Teacher Education at IALS), the decision was taken to reduce the amount of evaluation during TLE, on an experimental basis. Two reasons were put forward in support of this move:

1. participant feedback on all of the individual sessions to be included in the course was available from RAF returns, and this had been positive and stable over at least one summer (i.e. three courses); moreover, since there was no indication from enrolments that the new groups of participants would differ significantly in any way from previous groups, it could be predicted that reactions would fall into similar patterns;
2. time previously spent in collating and considering responses to weekly questionnaires (approximately 1½ hours per week x 9 over the summer) would be saved.

A related decision was that - to assess the effect of this change - tutors would be asked at the end of each course whether they would have preferred a continuation of the previous system, and participants whether they would have appreciated a formal opportunity to voice their reactions to sessions during the course.

6.3.2.2 Results

As usual, TLE ran three times in 1993, during periods known internally as Block 1, 2 and 3. During Block 2, in addition to the normal mixed-nationality group, there was an all-

Spanish group, which used the same basic timetable, materials and procedures. With the exception of the CD of the all-Spanish group, all CDs were familiar with the RAF.

The quotations that follow represent (a) oral comments of course participants during final day open discussions, summarised or noted down at the time by the Course Director (CD)/ another tutor/the researcher (b) a mixture of written and oral comment from CDs.

BLOCK 1

'12 out of 18 would have preferred weekly evaluation [of sessions, rather than end-of-course evaluation] because of memory problems'

(CD report)

'I would opt for weekly evaluation, on paper, of the "any comment ?" type rather than grading each session'

(tutor)

BLOCK 2: small all-Spanish group

'they felt that the end-of-course evaluation was sufficient and better than having weekly ones'

(CD report)

BLOCK 3: 18 on course; 13 present at evaluation discussion on final day. (CD = Course Director, T = Tutor, P = Participant.)

CD: "We used to have weekly evaluations, but we decided to stop them this year because we were getting very similar results. Would you have liked something like that, to have had a weekly review ?"

P1: "I think it's easier to evaluate [individual sessions] at the end of the week [than at the end of the course]. I'd forgotten some of the activities."

P2: "I don't know. I think I prefer a global view. The first week I was completely lost and if you'd asked me on the first Friday, I would probably have said, 'Yes, I agree with everything'".

[* * *]

CD (to whole group) "Would you have liked the opportunity to say something before now ?"

P3: "Yes, maybe. Maybe a discussion, a short one."

[* * *]

T: "Did the fact that we didn't ask for your reactions on a weekly basis give you the impression that we didn't care what you thought ?"

Ps: (unanimous 'No').

P4: "If we'd wanted to say something, we could have."

P1: "If you hadn't asked, I wouldn't have thought about it."

(verbatim notes taken by researcher)

At a meeting of CDs at the end of the series of courses in summer 1993, the consensus was that in the interests of in-course monitoring, there should be a formal opportunity for participants to express their views on a weekly basis.

6.3.2.4 Discussion

Judging from these findings, participant opinions are mixed as to the need for a weekly evaluation (of any kind). While the majority of participants in Block 1 expressed a preference for weekly evaluations, the all-Spanish group felt these to be unnecessary, and the few participants in Block 3 who voiced an opinion expressed somewhat conflicting views. The main reason advanced for weekly evaluations was 'memory', presumably prompted, as P1 in Block 3 implies, by the feeling that it is either impossible to comment on sessions one has forgotten or that opinions about sessions experienced two or three weeks earlier and perhaps only half-remembered lack validity. On the other hand - and this point seems to lie behind the comment of P2 in Block 3 - it may only be possible to make judgements about sessions once one is in a position to compare them with others or place them in the context of the overall plan.

It is perhaps worth making the point that some insight into participants' level of satisfaction would still be available even if formal weekly feedback were dispensed with. As noted in 6.2.3.1, tutorials at the end of Day 2 are designed to elicit early reactions and special wishes and, if continued, would pick up on any obvious signs of dissatisfaction. Moreover, with an average intake of 24, in IALS terms TLE is a relatively small course. An informal

atmosphere is deliberately cultivated and it is not unusual for comments to be made during coffee break or at lunch-time, when tutors often sit with participants (see Block 3, P4). Nevertheless, the feeling among CDs was - as indicated above - that staff needed to be able to monitor participant attitudes on a regular and formal basis.

6.3.2.5 Conclusions

The conclusion that the experiment leads to is rather different from the one originally envisaged, which related to the retention or otherwise of the RAF. It should be remembered that the RAF fulfilled two functions: (1) by providing feedback on participant reactions to each session of the course, it contributed to formative and summative programme evaluation (2) because monitoring took place on a weekly basis, action could be taken more or less immediately. While there was no support from staff for the continued use of the RAF in itself (i.e. in function 1), there is evidence that participants (though not unanimously) and tutors perceive some benefit in weekly evaluation at a more general level (function 2) and/or review (a pedagogic function). The letter at the end of Week 1 and the review at the end of Week 2, if the latter could be organised in such a way that all views were heard, would be appropriate for this purpose.

The general conclusion from the experiment is therefore that a weekly evaluation and/or review of some description ought to be built (back) into the course.

6.4 IMPLICATIONS OF THE CASE STUDY

6.4.1 Answers to the research questions

The study disproves the view that time constraints limit the evaluation possible on short courses to, for example, a weekly review meeting with participants and an end-of-course questionnaire. In the period covered by the study use was consistently made on TLE of additional measures: individual tutorials, weekly tutors' meetings, an end-of-course plenary discussion, self-evaluation and RAFs. Some use was also made of first-week letters. All of these proved feasible and useful for programme evaluation purposes, though the justification for individual tutorials as used on TLE is stronger in relation to individual contact and guidance. Moreover, there is no indication of participants feeling that too much evaluation was involved (*pace* Informant 4: 20) perhaps because this either took

place outside class time or was seen as integral to course process and objectives. The conclusion arrived at in 6.2.4.4 was that in addition to the evaluation procedures normally used on short courses of this kind consideration should be given to the use of RAFs, first-week letters and self-evaluation.

It is less easy to reach a conclusion concerning the second research question, i.e. whether the battery of evaluation procedures should be different or can be reduced as a programme becomes more established. The commonsense answer to this question is that more evaluation is likely to be necessary in the early stages of a programme, while it is at a developmental stage; as time goes by and the programme takes on a shape suggested by experience and feedback less evaluation is necessary.

This view is only partly supported by the 11-year study of TLE. The experiment described in 6.3, which addressed this question specifically in relation to the RAF, found no evidence to support its retention, its developmental function having been largely fulfilled. In other respects, however, there is no indication that less evaluation is needed, and this is hardly surprising. Tutors want to be in touch with participant reactions while a course is in progress and this is an argument for in-course monitoring instruments such as individual tutorials, whole-group review and first-week letters. Like tutors, programme-providers and designers will also want answers to questions which necessitate end-of-course (and/or post-course) evaluation: how the course compared with previous courses, which elements were successful, which were not, and how the latter might be modified. Moreover, if modifications are made on a subsequent occasion, further evaluation will be necessary to establish whether the resulting course is more successful in respect of the modified elements. Without evaluation of these kinds, programme design and delivery risk becoming random, unprincipled activities.

6.4.2 The need for review

The study has also shown that once evaluative procedures have become established there is also a danger that they become routinised. There seems little doubt that TLE continued to evolve and meet participants' expectations over the period 1982-92 because those who worked on it during these years took evaluation seriously, and the senior management of IALS responded to the CD reports by making development time available. The emphasis,

however, was on developing the course and on developing evaluation measures which would further this purpose rather than on assessing the efficacy of the evaluation measures themselves. The study has indicated that there is a need periodically to review the evaluative approach itself. Where a programme is only lightly evaluated, there is a case for considering whether more evaluation would be helpful; where a great deal of evaluation takes place, thought needs to be given to the cost of obtaining and processing the information, and the uses to which this information is put.

6.4.3 The need for documentation and standardisation

A final point suggested by the study is that documentary evidence is necessary if programme review is to take place. In this case, the absence of information (data, records of action taken) made it impossible to draw firm conclusions concerning the utility of specific instruments or the leverage exerted by CD reports. Comparison across courses also proved difficult because of inconsistencies in report formats (a point which has been addressed since the completion of the study). Careful records of evaluation (and subsequent action) are needed, in standardised form.

CHAPTER 7

CASE STUDY 3

A COURSE FOR ML PROBATIONERS AND RETURNERS

7.1 INTRODUCTION

This chapter extends the discussion of in-course evaluation for formative purposes.

The arguments for and against programme evaluation being carried out by course tutors were presented in Chapter 2 (2.5.3), a major consideration being the possibility that findings might be (or appear to be) biased by the relationship between participants and tutors or by a tutor-evaluator's personal investment in a positive outcome. Ways of guarding against such bias were also discussed in Chapter 2. In the field of language teacher education methods have included the use of researcher-evaluators conducting interviews and diary studies (Murphy O'Dwyer 1985; Morrow and Schocker 1993); the general education literature makes reference to the use of participant observers (e.g. Rudduck, J. 1981).

The findings of the survey reported in Chapter 4 provide only a little evidence to suggest that UK institutions see bias as a potential problem. While it is clear that in the context of certificated programmes value is attached to the reports of external examiners, the general lack of external input to non-certificated programmes leads one to wonder whether external examiners would always be used for certificated programmes if there were no compulsion to do so. The small number of exceptions include such measures as the involvement of staff from another department within the institution and the use of a course-external insider to lead evaluative whole-group discussions with participants. The use of

participant journals as an in-course monitoring measure was also mentioned by Informants 1 and 5. In both cases, however, the journals appear to be written for and analysed by course tutors rather than someone external to the course-team, and may therefore be subject to the same potential forms of bias as insider-conducted questionnaires or interviews.

This chapter sets out to explore the potential contribution to programme evaluation of participant journals written for an evaluator and of a non-participant observer. As Chapter 3 indicated, participant journals have been exploited for a variety of purposes, but little consideration has been given in the L2 teacher education literature to their use as an evaluation instrument. As far as observation is concerned, we find references in the field-specific literature to observation of participants teaching as part of a programme evaluation, but apart from Murphy-Dwyer (*op.cit.*) there appear to be no reported cases of non-participant observers being used in programme evaluation.

The research questions were therefore as follows:

1. *what are the* pros *and* cons *of participant journals as a programme evaluation instrument* ?

2. *what contribution can a non-participant observer make to programme evaluation* ?

7.2 THE STUDY

7.2.1 Introduction

The case study with which this chapter is concerned relates to a 'twilight' (early evening) course for Modern Language (ML) teachers in the Lothian Region of Scotland. Preliminary groundwork took the form of two lengthy interviews in October and November 1992 with Peter Wheeldon (PW), at that time Adviser for Languages, Lothian Region. The first interview was primarily designed to gather information on the nature of L2 INSET provision within the region and the means by which this provision is evaluated; the second dealt specifically with one particular form of provision: a twilight course, which had been run on two previous occasions, for probationary teachers and qualified

teachers wishing to return to teaching in Lothian. A summary of the interviews, which were recorded and transcribed, can be found in Appendix 7.1.

7.2.2 Evaluation questions

The discussions with PW revealed that apart from the rather unusual mix of participants, the programme for probationers and returners was unusual in allowing participants freedom to determine course content. The discussions also suggested that a careful evaluation of the programme by an outsider would be timely and welcome. The central question for PW as organiser was obviously the extent to which the programme was successful in meeting its stated aims (see 7.3.3, below) and participants' perceived needs, but given the characteristics of the programme noted above two further questions seemed worthy of study:

1. were the organiser's assumptions concerning the mix of probationers and returners (see Appendix 7.1, section 6.4) well-founded ?
2. did the participant-centred philosophy work ? if so, how ?

7.2.3 Overview of the case study

Section 7.3 provides brief details of the course. Sections 7.4 - 7.6 describe the evaluation measures (participant journals, questionnaire, plenary discussion) and the relation between these methods, touch on my role as non-participant observer and present the results obtained. Following a discussion (in section 7.8) of criteria for the evaluation of the course and the application of these, a number of recommendations are reported and an indication given of the extent to which these were subsequently implemented. The fiinal section considers the implications of the study.

7.3 BRIEF DESCRIPTION OF THE COURSE

7.3.1 Duration and intensity of course, location, participants, attendance

As indicated above, course number 93T130 was designed for ML teachers who were returning to teaching after some years (or a period outside Lothian) and probationers in their first or second year. Six meetings were arranged (16.15-18.00) between mid-

November 1993 and mid-January 1994 in an Education Centre in central Edinburgh. In session 5, shortly before Christmas, it was agreed to extend the course by a further three sessions and these took place between early February 1994 (one in the same centre and two, when the first centre was unavailable, in another centre. Of the 24 teachers enrolled, 5 were probationers, 11 returners currently teaching in schools and 8 would-be returners; 3 were native speakers of languages other than English. Attendance at each of the first 5 sessions (before Christmas) varied between 21 and 19; 18 attended at least 5 of the first 6 sessions. Numbers dropped a little for the February meetings.

7.3.2 Tutors

The course coordinator was Peter Wheeldon (PW). Since none of the tutors who had taught on the courses previously were available, PW was obliged to bring in a number of new tutors. These were: CS, a Lecturer in Modern Languages at Moray House Institute of Education, Heriot Watt University, who co-taught sessions 1-3 and 5-6 but was unavailable for the extension phase of sessions 7-9, and two school teachers, AM (2 sessions) and AB (1 session). PW handled the last two sessions alone.

7.3.3 Course aims

The explicit aims of the course are contained in the advance notice to schools, Circular no. 405/93, an extract from which appears in Appendix 7.3. These were:

1. Returners - updating on all recent developments in Modern Languages.
2. Probationers - identification of needs and their fulfilment.

In a private conversation during Week 3 of the course, PW amplified these aims in respect of the returners. 'Updating', he explained, had both a content dimension, e.g. familiarising participants with (changed) procedures in schools, and an affective dimension: the course was also intended to provide an opportunity for people to meet others in the same situation, give them ideas and techniques to hold on to in the initial stages and a metalanguage to discuss these new ideas, and as a result, should have a positive effect on returners' confidence.

7.3.4 Course content

Course content was negotiated on the basis of the menu of topics distributed in advance (Appendix 7.3). Participants who attended the first six sessions worked on some or all of the following:

Session 1	Communicative Methodology or Introduction to Standard Grade
Session 2	as above (same subgroups)
Session 3	Approaches to teaching S1 (Secondary School, Year 1)
Session 4	Learning Support Service
Session 5	as Session 1 (groups switched topics)
Session 6	Meeting the needs of the least able

(For the content of Sessions 7-9, see 7.8.2.)

7.3.5 Facilities and resources

For most of the sessions, participants were split between two minimally equipped rooms. In both centres, the larger room could accommodate all course members and the tutor seated around a block of tables. The seating arrangement was similar in the smaller room. In the last two sessions participants sat in groups to work on different tasks and the tutor moved between the groups.

7.4 EVALUATION PROCEDURES

7.4.1 Rationale for evaluation procedures

The evaluation questions necessitated data on participant reactions to the course and a record of course process.

What had been missing in the evaluation of the previous courses (see Appendix 7.1, section 6.4) was any quantitative data of the sort normally available as attendance records or collected by questionnaire. The decision was taken to keep a record of attendance and to collect quantitative data by means of a questionnaire. Although a standard questionnaire was used on other (typically one-day) courses, this was rejected as unsufficiently focussed

(see Appendix 7.2). Given the known differences between subgroups of participants it was felt desirable also to obtain qualitative data, and that **individual journals** might prove an economical and revealing source of such data. A **questionnaire** would then be used for the specific purpose of testing any patterns in this data for generalisability. Limitations in the questionnaire and any strongly-felt issues would be explored in **a plenary discussion** following the administration of the questionnaire. Each of these instruments is discussed further in subsequent sections.

Course process was also of interest. If completed by tutors as well as participants, the journals might reveal differences in role-related concerns or reactions, but since these might be so diverse as to be irreconcilable or uninterpretable, it would be helpful to maintain a third and more objective record of the sort that might be kept by a neutral observer. This person might also coordinate data collection from the other sources. The operationalisation of this idea is briefly described below.

7.4.2 My role

I attended the course as a non-participant observer. My identity and evaluative role were indicated to the group during the course introduction, and I was invited to say a little about how I proposed to evaluate the course. I used this opportunity to introduce the concept of individual journals in which participants should take notes and record their impressions of the course, and distributed notebooks for this purpose. The two core tutors, PW and CS, had already agreed to keep such journals.

Throughout the course I sat a foot or two behind and apart from participants to indicate that although I wished to be part of the group in a social sense, I did not intend to participate actively. This seemed to work well. I was included to the extent that participants passed me handouts (and even the register, sometimes), but I took no part in the discussion, was not asked to comment by the tutors and deliberately restrained myself from communicating non-verbally with the tutors. One consequence of non-participation was that I was free to observe and take notes on whatever struck me. During sessions when two rooms were used, I spent roughly half of the time with each group.

7.4.3 Getting to know people

Each week after Week 1, I deliberately arrived early to chat to participants informally. In Week 2, I spoke to four people; in Week 3 to a different three, and so on. Although I hoped to find out a little more about individuals and pick up incidental feedback, the basic aim was to offset any negative effects of my non-participation and status as official evaluator. Such effects, I felt, might manifest themselves in a reluctance to hand in the journals or careful editing of these at the writing stage.

7.5 JOURNALS

7.5.1 Introduction

The journals were seen primarily as an elicitation instrument, a means of gaining some insight into the reactions to the course of participants and tutors, reactions which could be checked against my own field notes. Individual responses, suitably reformulated, could then be used as the basis for a questionnaire to be administered in Session 6, the final session of the course, according to the original programme.

Careful prior consideration was given to such questions as:

- the format of the journal;
- how the concept of a journal, and evaluation related to this, would be presented to participants and tutors;
- the staging of this presentation;
- the use to be made of the journals.

Sections 7.5.2 - 7.5.5 record the factors explicitly considered and the steps taken.

7.5.2 Journal format

Following Murphy-O'Dwyer (1985), I decided to give participants a notebook in which to keep the journal rather than ask them to take notes on whatever they wished. This, I felt, would lead them to take the evaluation exercise more seriously.

Various criteria influenced the selection of the notebook: convenient size (8" x 5"), number of pages (80), lined, spiral binding (easy to remove anything writer did not wish researcher to see), cost (at 35p a real bargain !).

7.5.3 Presentation of the journal as an evaluation measure

Week 1: at this stage, I merely pointed out the advantage of journal-keeping. I suggested that participants write during sessions and between sessions, but did not not indicate what kinds of thing to write (i.e. I adopted a basically non-interventionist policy deliberately, with a view to observing subsequently any differences between non-directed journal-keeping and the alternative). One person asked rather anxiously how much they would be expected to write between sessions; PW suggested two or three sentences.

Week 2: I asked how many people had added to their entries in the intervening week. This was intended as a gentle reminder of the importance attached to keeping the journal and as a prompt to those who needed one to write something between sessions. Since only six or seven signified that they had written something since the last session, the reminder was probably timely.

Week 3: I had prepared a handout (Appendix 7.4) on pink paper (to distinguish it from the course handouts, which were on white). The hand-out drew attention to some of the advantages of journal-keeping for the participant, reinforcing and extending what had been said in Week 1; it also indicated that the journals would be taken in and how they would be used. Confidentiality was stressed. These points were also made orally at the end of the session, when the handout was distributed.

Week 4: I checked at the end of the session that everyone had received a copy of the pink sheet the previous week and reminded them that the journals would be collected at the end of the following session.

Week 5: I collected the journals at the end of this session. Following a preliminary analysis, I photocopied relevant sections so that the journals could be returned in Week 6.

7.5.4 Other factors affecting the study

One particularly **important** factor in encouraging participants to take the journal seriously was the validation provided by PW, as course coordinator. In introducing me at the first session, he emphasised that he had asked me in, as an experienced professional, to help him evaluate the course. And when I asked, at the beginning of the second session, how many participants had made entries in the course of the week, PW followed this up by asking if they had understood the sub-text of this question, and encouraged them to cooperate.

PW's flexibility and cooperativeness had been evident in our preliminary discussions, but his active support during the study was also reassuring. In the third session, we sat out for a few minutes while he looked through the hand-out I had prepared (Appendix 7.4); Peter gave full support to what I had in mind, even though this would mean allocating approximately one third of the session in Week 6 to questionnaire-completion and an evaluation discussion.

7.5.5 Use made of the journals

Analysis of the journals provided two kinds of input to the evaluation of the course. First, it indicated individual reactions and patterns of reaction which informed the evaluation instruments used in Week 6 (the questionnaire and plenary discussion). Less directly, it provided some evidence of ways in which participants had been stimulated to reflect on - and act on - ideas presented during the course; however, such evidence was suggestive rather than conclusive, since a number of participants preferred to use their own pads for notes and many wrote directly on to hand-outs when these constituted session notes (e.g. Week 4).

The idea for a questionnaire deriving from a preliminary study of the target group is an adaptation of an idea of Norman Evans (1987). In a case study included in Powney and Watts (1987), Evans describes his approach in the late 1970s to the evaluation of the in-service B.Ed. degree then offered by some sixty institutions. He began by looking at documentary evidence from all the programmes and established a set of categories (e.g. use made of participants' experience; workload; course structure) within which to try to capture the reactions of those involved. The categories were piloted, revised, and then

used as the basis for conversations with participants and tutors from six carefully selected programmes. Some of the conversations were individual, others were in groups; all were audiorecorded and transcribed. Evans then searched through the conversations for what he calls 'significant statements' (Powney and Watts 1987: 89) under each of the categories. 47 statements were subsequently circulated prior to a second round of conversations with participants and tutors from the same sample. 'What this amounted to,' Evans comments, 'was an attempt to get as close as possible to the evidence about In-Service B.Ed. programmes which seemed most important to those who were teaching it and studying it' (*ibid*).

7.5.6 How the journals were completed

There was wide divergence in the approaches to the writing of the journals, as one might expect, given the (deliberate) absence of initial guidance. The suggestion was that participants should use the notebooks for both note-taking during sessions and reflection on the sessions. As indicated above, some participants took no notes in the notebooks, preferring to write on hand-outs or in A4 pads (one mentioned that the latter system allowed her to file her notes more easily). Approximately half the group failed to respond to the request for comments, despite the repeated, carefully staged attempts to encourage this (see 7.5.3, above).

Among those, in addition to the two tutors, who did offer comments - two probationers, four returners (teaching) and four non-teaching returners - writing patterns also varied, from interspersed in-session comments, mainly of a very abbreviated nature, to the regular post-session comment (the majority preference) to longer retrospective pieces, after three or four sessions. The more measured comments varied in length from approximately 60 words (a little less than half a page) to several pages (described by the writer as 'a diatribe'). One returner, explaining the absence of any evaluative comments in her notebook, wrote that she would be better able to evaluate the course when she was back in the classroom.

7.5.7 Content analysis

7.5.7.1 Analysis strategy

No effort was made to assess the perceived value of the sessions on the basis of the descriptive notes taken, one possible indicator of what participants judged to be valuable/worth noting. As indicated above, the notebooks were not used uniformly by participants; moreover, the group composition was such that an analysis of this kind would have been likely to provide evidence only of the wide differences in experience, training and current situation.

A better guide to perceived relevance was thought to be (a) explicit evidence of immediate follow up among those participants in schools (b) evaluative comments.

It should be emphasised that the evidence which is reviewed below relates only to participants' reactions to the first five sessions of the course. In the original plan, this would have been a reasonable basis for evaluation; given the fact that the course eventually ran for nine weeks, the results should obviously be treated with some caution.

For ease of reference, the journal keepers whose journals did include evaluative comment were assigned code numbers and a prefix (P being used for probationers, RT for returners who were currently teaching, R for returners who had not yet returned to the classroom, and T for tutors).

7.5.7.2 Evidence of follow up

The journals yield very little evidence of ideas or content being applied in participants' schools - although the absence of any evidence to this effect should not be taken to mean that this did not happen.

P1, following Week 3, re-examined her own tasks and worksheets with an eye to differentiation; she also spoke to staff regarding LSS support.

RT3 was encouraged by Week 1 to create and try out material which proved 'popular'; after Week 3 she spoke to LSS staff.

RT5 frequently questioned and explored the ideas presented: e.g. 'Is Foundation "a badge of failure" ? - still a key to something further'. After the LSS session, she was led to consider whether she was herself fulfilling a kind of LSS role.

7.5.7.3 Evaluative comments

A rough categorisation and quantification of participants' comments can be found in Appendix 7.5. In this section the focus is rather on triangulation, the comparison of events from three points of view: that of the participants, that of the tutors, and that of the observer. Of particular interest from an evaluative perspective are the points on which there is concurrence of opinion; any striking differences may also prove illuminating, of course.

The bulk of participants' comments relate to Session 1. These range from positive statements concerning the perceived value of the session (and what contributed to this feeling) through doubts about the mixed nature of the group and uncertainty regarding other participants' backgrounds and needs to negative comments on classroom management ('disorganised') and the effects of this ('confusion'). Subsequent positive entries relate mainly to value ('practical', 'useful') and interest; negative points, which are few in number, express frustration with organisation (e.g. lack of structure and directedness, not being able to continue with a topic started previously) and with support materials; there is also comment on the relevance of the session on Learning Support Services.

There are a number of points of similarity between the tutors' journal entries and those of participants. This is especially true of T1. In relation to Session 1, T1 expresses unease about the way in which the menu was used and priorities and aims decided. He wonders whether name tags might have been used, and notes that the mixed nature of the group needs to be addressed, a theme which recurs in each of his subsequent entries and culminates, in Week 5, in the conclusion that probationers and returners need separate courses. Following Session 2, unconsciously echoing one of the participants, he wonders if

the returners (teaching) are 'railroading' the non-teaching participants, and following Session 3 comments positively on the contribution of the questions asked by members of the group. The decision to spend the whole of Session 3 on S1 teaching clearly took him by surprise; he observes that the group he had been teaching the previous two weeks had been keen to continue. T1 was present in sessions 1-3 and 5-6 only.

T2's preoccupations appear to be rather different. Administrative details figure, in the form of references to coffee cups and the crêche organised for a desperate mother, and he is concerned about time: the time needed to deal adequately with concepts (Sessions 1, 2) and the time actually taken up by activities; he is also anxious about overrunning (Sessions 1 and 4). Pedagogical aspects of sessions interest him: he refers in broad terms to his own approach to the content of sessions and makes observations about group reactions (all sessions) and such details as the difficulty of a task and the contribution of examples to participants' understanding. T1 is similarly concerned about group reactions, but more in relation to the differences between the subgroups: for instance, he observes that although most of his group participate, some obviously prefer to listen (Session 2) and that the session in which he split non-teaching returners and probationers/teaching returners for small group discussion seemed to work satisfactorily, a point on which one of the probationers in his group concurred.

My own notes, as observer, are naturally much more detailed than those of either the participants or the tutors. During sessions I tended simply to describe and register questions for later consideration. After the sessions I wrote up my notes and drew tentative conclusions. The points on which my own observations tallied with those of participants and at least one of the tutors are listed below:

1. menu seemed to be used in rather perfunctory way; topics decided and groups formed with very little discussion
2. no negotiation of syllabus for course as a whole (i.e. long-term view)
3. lack of formal opportunity for participant self-introductions
4. differences in participant starting-points
5. core tutors (CS) make good use of their own practical experience

6. possible frustration resulting from apparent change of plan in session 3 (i.e. to devote the whole session to approaches to S1 teaching rather than spending part of the time on topics already begun)
7. session 4 (on Learning Support Services) illuminating, but ...
8. task set by AB difficult for those not in schools

In some cases, my own observations matched those of participants only. These related to:

9. uncertainty concerning starting time of sessions
10. lack of discussion of course objectives
11. relaxed atmosphere

My notes also include a number of points mentioned by neither participants nor tutors. They are included here in order to complete the record, as it were, but also because such insights are more easily obtained by a detached observer:

12. the need to discuss the concept of learner-centredness as it applies in teaching and as it affects participants on the course
13. systematic way in which tutors elicited ideas from participants
14. clear evidence of tutors' ability to relate specific topic to broader educational concerns

Many of these points were incorporated in the Questionnaire (see below and Appendix 7.6); points 12-14 are picked up under 7.8.3, below.

7.6 THE QUESTIONNAIRE

7.6.1 Design of the questionnaire

Any points made in the journals of an other than descriptive nature were considered for inclusion as 'significant statements' in the questionnaire (Appendix 7.6), which was intended to convert such qualitative data into quantitative data. Taking into account such factors as saliency (repeated mention in participant journals), congruence with tutors'

journals and observer's notes, course aims, tutors' expressed concerns and time available for completion and layout, a set of 19 statements was drawn up. A 1-5 scale was adopted, labelled at the extremes in an attempt to capture shades of reaction; and participants were asked to give their names and indicate whether they were probationers or (teaching) returners. An open slot was included to allow for any open-ended wishes to be expressed.

7.6.2 Administration of the questionnaire

Session 6 opened with a few announcements from PW. I then introduced the questionnaire briefly. As expected, it took about five minutes to complete. The questionnaires were then collected and subsequently analysed. Neither these nor the journals were shown to the course coordinator.

7.6.3 Plenary discussion

Following the collection of the questionnaires, I invited any comments on points touched on in the questionnaire or on any other matters the participants wished to raise. The resulting discussion, which was recorded using a Tandberg Audio Tutor and a Sony TCS 2000 (to be on the safe side) and which took approximately 15 minutes, is summarised in 7.7.3, below.

7.6.4 Results of the questionnaire: scale questions

Of 24 possible respondents, 21 completed the questionnaire (18 in Week 6, 3 in Week 8). Patterns of individual response to the scale questions were as follows:

- 4 used all 5 points on the scale;
- 9 used 4 points (4 avoided point 5; 2 did not use point 1; 2 did not use point 4);
- 5 used 3 points (of whom 4 used only 1-3);
- 3 used 2 points (2 used points 1-2 only; 1 used points 1 and 3).

From these figures, we might conclude that that although the majority used the scale selectively, a number may have felt somewhat constrained. During the subsequent discussion, one person pointed out in reference to the journal that he preferred not to criticise. Since participants were asked to put their names on the returns (to allow for

cross-referencing with the journals) others may have also preferred not to appear critical (for the reasons discussed in Chapter 3).

Table 7.1, below, shows the responses to the 19 statements by scale point and by subgroup: P(robationers) n = 4; R(eturners) T(eaching) n = 9; R(eturners) n = 8. Totals (Tot.) for each scale point are also indicated, as are instances of 'no response' (nr).

Table 7.1: Responses to statements by scale point and subgroup

	1		2		3		4		5		nr
SUBGROUP	P RT R	Total	P RT R	Total	P RT R	Total	P RT R	Total	P RT R	Total	
Statement 1	0 1 3	4	0 4 3	7	2 4 1	7	1 0 0	1	1 0 1	2	0
Statement 2	2 2 4	8	1 6 1	8	0 0 0	0	0 0 2	2	1 0 1	2	1
Statement 3	2 5 6	13	1 4 2	7	1 0 0	1	0 0 0	0	0 0 0	0	0
Statement 4	0 2 1	3	2 1 3	6	0 4 2	6	2 1 1	4	0 1 1	2	0
Statement 5	1 6 4	11	3 2 2	7	0 1 1	2	0 0 0	0	0 0 0	0	1
Statement 6	0 1 1	2	1 2 1	4	1 4 3	8	1 1 1	3	1 1 1	3	1
Statement 7	0 2 3	5	0 3 4	7	1 4 1	6	2 0 0	2	1 0 0	1	0
Statement 8	0 1 1	2	1 0 2	3	2 4 1	7	1 2 2	5	0 1 1	2	2
Statement 9	2 2 2	6	1 4 4	9	0 2 1	3	0 1 1	2	0 0 0	0	1
Statement 10	2 3 4	9	0 5 1	6	1 1 0	2	0 0 0	0	0 0 1	1	3
Statement 11	0 0 3	3	0 3 2	5	2 2 1	5	2 1 1	4	0 2 0	2	2
Statement 12	0 1 3	4	1 6 4	11	2 1 1	4	1 1 0	2	0 0 0	0	0
Statement 13	1 3 5	9	2 4 3	9	1 2 0	3	0 0 0	0	0 0 0	0	0
Statement 14	1 2 3	6	1 3 1	5	2 2 1	5	0 2 0	2	0 0 1	1	2
Statement 15	1 3 2	6	0 2 4	6	2 2 2	6	1 0 0	1	0 2 0	2	0
Statement 16	0 1 1	2	2 1 1	4	0 2 5	7	1 3 0	4	1 1 0	2	2
Statement 17	1 5 4	10	2 3 4	9	0 1 0	1	0 0 0	0	0 0 0	0	1
Statement 18	1 2 2	5	2 3 3	8	0 2 2	4	0 1 0	1	0 0 0	0	3
Statement 19	0 3 2	5	0 2 5	7	1 3 0	4	2 0 0	2	0 0 1	1	2

At first sight, these results are rather inconclusive. Not only are there no 100% concurrences, there is not even agreement within the subgroups. The strongest response was in relation to the menu of topics (statement 3 in the questionnaire), 13 of the 21 respondents selecting scale point 1 ('strongly agree') for this statement. The idea of breaking into two groups (statement 5) attracted 11 responses at scale point 1; good use of time during sessions (statement 12) also attracted 11 responses but at the level of scale point 2 ('agree').

By combining scale points 1 and 2, however, we arrive at some indication of tendencies. For instance, in response to statement 3, already referred to, only one respondent chose a scale point other than 1 or 2 (and this was 3, the central point on the scale). This and other strongly 'positive indications' are listed below, with combined figures for scales 1 and 2 shown in brackets:

Positive indications

2. good idea to let participants determine course content (16; 1 nr.)
3. good idea to have menu of topics (20)
5. good idea to break into two groups (18; 1 nr.)
13. (there were) opportunities to learn from experienced tutors (18)
17. sessions finished punctually (19; 1 nr.)

By the same method, by combining scale points 4 (disagree) and 5 (strongly disagree) we can arrive at a number of negative indications, although these are much weaker. The relevant statements are shown below:

Negative indications

4. way topics were selected from menu was well organised (6)
6. way groups were formed was well organised (6; 1 nr.)
8. topics have been systematically ordered (7; 2 nr.)
11. sessions started punctually (6; 2 nr.)

Bearing in mind the relative reluctance of a significant minority of the group to use the negative scale points and the largish numbers (5-7, or up to a third of the respondents)

sitting on the fence of scale point 3 in respect of the latter statements, there may be reason to suppose that dissatisfaction with these features of the course was stronger than the raw figures suggest.

In general the nine statements cited above relate to the participant-driven philosophy underlying the course (statements 2, 3) which met with general approval in principle, and classroom management (4, 5, 6) or organisation/administration (8, 11, 17), some aspects of which attracted criticism. Statement 13, which concerned opportunities offered by the course to learn from experienced tutors, is presumably one element in participants' overall judgement of the value of the course.

In this connection, responses to other statements are also of interest. Figures in brackets indicate the number of responses in scale points 1 and 2. Only statements attracting positive responses from at least half of the respondents are included.

7. course content has been relevant to my needs (12)
9. session on LSS useful (15)
12. time during sessions has been well used (15)
14. opportunities to learn from other participants (11)
15. opportunities to discuss matters of mutual concern (12)
19. now feel more confident (12)

Analysis by subgroup is a little more revealing. Returners, especially non-teaching returners, appear to feel much more positively about the relevance of the course (statement 7) than probationers, a point reinforced by the comment by two probationers in response to statement 20, the open item on the sheet, that they would have preferred a course which provided specifically for the needs of probationers. The single session on LSS, on the other hand, met with a generally positive reaction, the doubtful reactions being confined to returners (both subgroups) and possibly being explicable by lack of contact with the work of LSS teachers.

The perceived value of a course may lie not simply in the relevance of the topics treated but also in the opportunities it provides for vicarious learning (statements 13, 14) and for

discussion of matters of concern (statement 15). While there is evident appreciation of the tutors (no explicitly negative responses), a minority, spread across the three subgroups, either failed to value the opportunities provided for discussion or learning from other participants or would have liked more such opportunities.

As a confidence-boost (statement 19), the course appears to have worked well for the majority of returners, including those currently teaching; although probationers responded less positively, it may perhaps be assumed that for them this was not a particularly relevant issue.

A small number of statements are not covered by the above comments. These are grouped below:

10. liked meeting each week in the staffroom over coffee
1. objectives of the course were clear
18. good use made of support materials (handouts, etc)

Statement 10 (meeting in coffee room) should be seen in relation to statement 11 (sessions started punctually), which attracted six negative responses. There appears to have been some misunderstanding about the starting-time of sessions, which may have arisen from the time stated on advance course materials and notices. In practice, '4.15' turned out to be '4.15 for 4.30'. Although this was a deliberate strategy on the part of the course organiser to allow time for teachers from far-flung schools to reach the centre and an opportunity for others to get to know each other, it was stated explicitly for the first time only in the letter sent out between sessions 6 and 7. For those who could only arrive at 4.30 (or later), there was no opportunity to get to know others informally (a point touched on in reference to the absence of participant introductions in the first session); and for those concerned about a prompt start, the lingering over coffee may have been rather irritating. A similar point may be made in reference to statement 1 (clarity of objectives). The objectives were stated in the information circulated to participants in advance of the course (see Appendix 7.3), but were not discussed during the first session. In both cases (statements 11 and 1), negative responses may have been prompted by an expectation of greater explicitness on the part of the course organiser.

Statement 18 (use of support materials) attracted a rather lukewarm response. Individual journal entries suggest that at least one participant would have liked to see past Standard Grade papers earlier than session 8; wishes were also expressed in respect of differentiated materials and speaking tasks/tests selected by course tutors.

7.6.5 Individual Comments (completion of "I would have liked ...")

The final question invited participants to complete the statement: 'I would have liked ...'. All responses are listed below. Numbers in brackets indicate grouping of similar comments.

Probationer group:

- course specific to probationers' needs (2) (or timetable of subjects to be covered in advance)
- advance elicitation of preferences
- more in-depth study of the given programme
- more information on differentiation and motivating the unmotivated (this observation was made before PW's sessions on 18 January and sessions 7-9, which offered opportunities for discussion of this theme and related practical work)
- more work on the skill of writing

Returners (teaching):

- more time initially on how group would organise itself and course content, etc.
- more structure
- more definite change of group after first few sessions on same topics
- (would have liked) to attend all meetings: don't know how topics were selected
- opportunity to observe LSS teacher in action
- to have a basic understanding of the workings of LSS before considering its role in Modern Languages
- practical session on use of PALE units, video camera, etc
- more methodology
- more exam work
- discussion of revised Higher exam

- more time
- [I found all the handouts useful]

Returners (not teaching):

- more directed programme of topics
- information about course materials
- more exchange of experience with other participants
- [no particular wishes; course has been very helpful]

To some extent, these comments speak for themselves. They provide further evidence that some participants would have liked more direction or at least more time for the group to decide on the direction it wished to take. They also indicate the range of perceived needs and the difficulty of catering for all of these within a limited number of sessions.

7.7 PLENARY DISCUSSION

Some of the points referred to above also arose during the short discussion which followed the completion of the questionnaire. There were comments, for example, on the heterogeneity of the class (needs of probationers, who have experience of Standard Grade, for instance, being different from those of returners; needs of returners in schools, who have some understanding of the nature of Standard Grade, being different from those hoping to return). Two participants commented that they had seen some of the materials before, in a session at Moray House and at an in-service day. One returner felt that she would have got less from the course if the class had not contained people with recent classroom experience, while recognising that from this point of view the course may have been less useful for them.

Specific suggestions included:

1. a preliminary introductory evening for returners (resumé of exam system);
2. exchange of personal information on the first evening to raise awareness of commonality of backgrounds and needs, and therefore inform topic choices, grouping and the approach to particular topics;

3. more direction from those (i.e. the tutors) who could predict the needs of specific groups.

All these points had previously surfaced in journals.

New points to meet with support were the idea for a predominantly practical session concerned with e.g. PALE units and the use of videocameras, and for a window on classrooms in the form of videorecordings. The discussion also led one participant to express the wish (previously voiced in her journal) to sit in on classes, a wish which, it was made clear, could easily be satisfied.

7.8 CRITERIA FOR EVALUATION

7.8.1 Participant satisfaction

In themselves the journals do not provide a basis for generalisation about participant satisfaction. The questionnaire returns, on the other hand, despite the reservations expressed in section 7.7 about certain features of the course, give the impression that the majority of participants found the the first five weeks of the course (the period covered by the journals and questionnaire) to be worthwhile.

This impression is borne out by attendance figures. As can be seen from Figure 7.1, below, with the exception of two probationers who opted out early, partly because of school commitments, there were relatively few absences during the first six weeks. 12 attended each of the first six sessions and six more missed only one session. Most of these absences were explained in letters or telephone calls to PW.

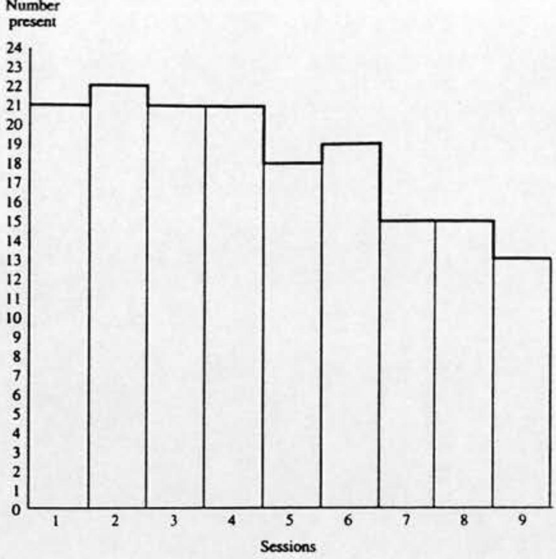


Figure 7.1: Attendance figures, by subgroup

It is true that would-be returners particularly may have felt that attendance was expected or at least advisable if they wished to be kept informed of job openings, but apart from one journal entry which questioned the status of the course, there is no evidence to suggest that such considerations affected either attendance or questionnaire returns.

Attendance figures fell by approximately 25% in the extension phase of the course, seven attending no sessions after Week 6. This may be explained by reference to prior commitments, fatigue or the feeling that the most immediate needs had been satisfied.

7.8.2 Fulfilment of programme aims

The stated aims of the programme were (a) to update returners and (b) to analyse the needs of probationers and cater for these needs.

One problem with a participant-driven course is that it can only work efficiently if participants are aware of their needs and willing to state these. Now while it would be reasonable to expect probationers and returners already teaching to be aware of their needs, would-be returners are not in that position. In a sense, they need to be told what their needs are. Moreover, participants' willingness to state their needs will depend on a number of factors: for instance, whether they are aware that their needs are shared by others; whether they feel the elicitation of their needs is a genuine commitment to meeting them; whether they feel confident enough to speak out in front of a strange group, and so on. Some preliminary sharing of personal information, which would enable individuals to locate themselves in relation to others in the group, could obviously help to encourage the expression of needs.

Participants who attended the first six sessions spent time on the following topics:

- communicative methodology: principles and (some ideas for) practice
- the structure of the new Standard Grade exam
- assessment of Standard Grade speaking
- approaches to S1 teaching
- the role of Learning Support Services
- motivating the unmotivated (the psychology of the least able learner)

Sessions 7-9 focussed on the following, the last three items being options from which participants had to choose:

- dealing with the unmotivated (disciplinary procedures, etc)
- marking Standard Grade written exams
- the structure of the Revised Higher exam
- materials for the least able learners

Relevant though these topics are, they constitute a rather patchy coverage of developments in Modern Language teaching, one noticeable omission - raised during the evaluation discussion in session 6 - being the use of technology. One participant wrote in her journal:

> How do we know we are actually going to cover what we most want to do ?
> Perhaps we could have drawn up a programme w. specific titles eg
>
> 1. ! AIMS? How to be achieved? 1. S1 + S2
> 2. S3 + S4
> 3. Speaking skills - approaches to
> 4. Listening
> 5. Reading/writing skills
> 6. Dictionary work
> 7. The slow learner/the "good pupil"/the disruptive pupil
> 8. Homework/Resources/How organize classroom"Use" of Assistants
> 9. On to Higher things etc

Such a course would, in fact, have been in certain respects less satisfactory than the one offered, which clearly did deal with the need for information (e.g. about approaches to S1 and the structure of Standard Grade and Revised Higher) as well as offering opportunities for practical work (e.g. on assessment of spoken and written work) and exchanges of

experience and ideas about dealing with slow and disruptive learners. What the list does point to, however, is that there inevitably remain a number of unsatisfied wishes/needs, and the necessity for a more thorough discussion of aims in relation to these.

The issue of coverage notwithstanding, 12 of the 16 returners to answer the question about confidence in the questionnaire administered in session 6 stated that they felt more confident as a result of the course; and this proportion might reasonably be expected to be higher the longer the course continued.

The subgroup which were apparently least well catered for were the probationers. As a minority group, they had a minority voice when it came to group decisions, and their responses to the questionnaire were significantly more lukewarm than those of returners. Moreover, as noted above, two felt that a separate course for probationers would have been preferable.

On the basis of this evidence, it appears clear that the organiser's views concerning the desirability of mixing probationers and returners were not well-founded.

7.8.3 Other effects

Informal comments suggest that the content and format of sessions 8 and 9 was well received. In these sessions, groups of participants worked cooperatively on different practical tasks with the tutor spending time with each group in turn. One beneficial effect of these sessions was that participants saw a philosophy (of learner-centred, task-based learning) in action and having experienced it as something positive, would be more likely to try it out with their own learners.

These final sessions also illustrated a feature of all the sessions: the striking commitment of all four tutors. Their combination of realism and enthusiasm and their belief in the value of what they are doing and their role as educators in a broad sense (and not simply subject instructors) had - in my perception at least - the power of inspiration. The experience of working with such tutors and the knowledge that, as PW made clear at the end of the last session (and demonstrated in tangible ways throughout), participants now had a direct link

with a source of practical support, may ultimately prove to have been the most lasting effects of the course.

The conclusions drawn from this evidence formed the basis for recommendations to PW. These, and a note on their implementation, can be found in the next section.

7.9 RECOMMENDATIONS AND IMPLEMENTATION

7.9.1 Recommendations

Recommendations, addressed to PW as programme organiser, were as follows:

> It would be difficult (impossible, in fact) to cater fully for the range of needs and wishes expressed by the group. However, given existing resources, the following suggestions might go some way towards meeting the needs of all three sub-groups (probationers, returners teaching and would-be returners) within a common framework.
>
> 1. Clarify the status of the course (wholly optional ? required for returners ?).
>
> 2. In determining dates for the course, take into account predictable demands on those participants who are teaching (e.g. marking towards the end of term, rehearsals for Christmas events). Consider blocks of time (3 hours) for certain sessions if this is feasible.
>
> 3. Distinguish between probationers and returners (and between returners who have started teaching and those who have not). This might be done in the following ways:
>
> (a) Establish the needs of the two returner sub-groups under topic heads. Decide which of these are pre-requisites for teaching in Lothian schools, order these in a logical sequence, and make them the basis of a core course. A distinction can be made between returners currently teaching and those not yet in schools by providing differentiation in, for example, the second half of each session and/or by offering one or more introductory sessions aimed specifically at the latter group and covering major developments in the last ten years or so, i.e. an 'Orientation to teaching Modern Languages in Lothian schools in the 1990s'. One of the core sessions, ideally including a hands-on component, should focus on the use of teaching aids which may be new to returners (e.g. various arrangements for listening and recording; videocameras). For returners not currently teaching, visits to schools should be arranged. The emphasis of this part of the course should be on: (a) raising awareness of the nature of the changes in the last decade and the rationale underlying these (b) establishing a common language for the discussion of concepts and techniques (c) providing a forum in which common concerns can be aired and discussed.
>
> (b) Towards the end of this first phase of the course, circulate a menu of possible topics for a second phase to probationers and returners already teaching. On the basis of expressed wishes, draw up a list of topics to be offered in this extension phase (of perhaps four more, possibly longer sessions - see 2, above). These should also be made available to the non-teaching returners. A range of interests could be covered by the method employed in 1993-94, i.e. by split-group sessions, participants opting in on a weekly basis. The

emphasis of this phase should be on a cooperative approach to common problems such as devising and applying marking schemes and the preparation of differentiated materials.

4. Ensure that there are is an opportunity in the first session for participants to get to know a little about each other on a personal and professional level.

5. Be explicit about the following:
- the objectives of the course (and its separate phases if recommendation 3 is accepted)
- expectations concerning start and finish time, and absence
- the learner/participant-centred philosophy underlying the course

6. Retain those features felt to be strengths of the course in 1993-94: e.g the participant-centred philosophy with built-in choice and differentiated grouping; the use of a number of experienced tutors. (Note that recommendations 1-5 are based on the main points to have come out out of the questionnaire, plenary discussion and journals. For other, more specific points, see the Appendix [here reproduced as App. 7.4.])

7.9.2 Implementation of recommendations

The programme continued in 1994-95. At a meeting in early 1995, PW confirmed that the major recommendations had been accepted, in so far as this was feasible. Constraints in the form of his own commitments and the availability of suitable premises had meant that an earlier start had not been possible (Recommendation 2), but steps had been taken to implement Recommendations 1, 3 and 5. The confusion that had arisen in relation to the optionality or otherwise of the course in 1993-94 (Recommendation 1) had apparently been due to the fact that some participants had been directed to the course by the Personnel section in the Education department; PW had written to senior staff in this department pointing out that the course was intended to be optional both as a whole and in its separate parts. Following recommendation 3, separate courses had been offered to probationers and returners and the course for returners had begun with a number of sessions the content of which had been determined by the Adviser; a menu system was then adopted, as previously. Recommendation 5 was implemented following session 1 of the 1994-95 course and at the point when participants were offered a choice.

While accepting in principle the case for recommendation 4, PW pointed to the difficulty of achieving a balance between climate-setting and content. His own experience of courses in which the balance had, in his view, erred on the side of the former had, he admitted, perhaps led him to go a little too far in the other direction. He nevertheless felt that some responsibility for self-introduction should lie with participants; after all, they would need to be able to establish themselves in the community of a school.

7.10 IMPLICATIONS OF THE CASE STUDY

7.10.1 Introduction

The courses prior to Course 93T130 had not been systematically evaluated. Participants had even resisted PW's attempt to get them to complete a questionnaire. One option for Course 93T130 would therefore have been to conduct an open review session at the end of the course or at the end of each session. In itself this would not necessarily have been more consuming of class time than the evaluation approach adopted, but it would have had certain disadvantages in other respects. Unless preceded by an opportunity for reflection and discussion among participants, following for instance the procedure described in Appendix 4.6 by Informant 1 (1: 4), for which class time would have been a requirement given the constraints on participants meeting, the discussion might have been dominated by a few individuals voicing views which were not necessarily representative.

The rationale for the evaluation instruments selected was that these would give a more complete picture of course process and participant reactions both to this and to course content. Of particular interest from the research perspective was the extent to which these instruments yielded information - for course development purposes - that would not otherwise have been available. This section considers the contribution of the evaluation instruments previously described but also touches on other features of the evaluation design, the use of a non-participant observer and triangulation.

7.10.2 The journals

The journals contributed to data collection in two ways: they revealed concerns at an individual level (7.5.7.3) which were explored subsequently by means of the questionnaire and post-questionnaire plenary discussion, and they provided positive evidence, however slight, of take-up by participants - and the nature of this take-up (7.5.7.2). At a more general level, they also offered a private, relatively convenient medium of communication between individual participants and the evaluator (a form of contact that would be equally appropriate for tutors wishing to evaluate their own teaching - Porter *et al* 1990). From a research perspective the first of these functions is the most relevant.

However, the fact that fewer than half of the potential informants used the journal in the way that was intended merits some comment. Judging from the questionnaire returns, a minority of the participants were unwilling to make negative comments about the course (although logically this should not have prevented them from making positive comments). It is also possible that the way in which the journal was presented was too oblique: that some participants did not realise what was expected of them. A more plausible reason may be that once the sessions were over these students, like many others, closed their notebooks until the following session. If this is indeed the case, it would argue for a period of, say, 10 minutes to be allowed at the end of each session for reflection in writing on the session.

Like any spontaneous data, journal data poses problems of analysis, particularly in relation to the comparability of results. These problems might be reduced if a more directive approach were adopted, including an indication of the topics on which comment is expected and examples of the kinds of comment that would be appropriate. There is a danger, however, that this would be to slant the journal too obviously towards evaluation, with a consequent effect on participant motivation. An alternative would be to ask participants to organise their entries in two sections: one being a review of what had been done and the other a comment on that (which might include reference to follow-up). This would permit comparison both of the notes on content (a possible way of assessing group and individual understanding of salient points) and reactions to course content and course processes.

7.10.3 The questionnaire

When the study was planned, it was assumed that the course would last for only six weeks (although previous courses had continued for longer in response to participant demand). The decision to locate the questionnaire in the sixth session was therefore deliberate: this allowed for at least four entries in the journals (which were collected at the end of Session 5) and administering it at the beginning of that session ensured that its completion would not be a hurried affair.

What distinguished the questionnaire in this study from the kinds of questionnaire normally administered at or towards the end of a course is that rather than merely reflecting the

concerns of course tutors it was based on input from three sources: the participants, the tutors and the observer. Its statement and scale format allowed for the economical collection of reactions to a fairly wide range of issues and the request for respondents to identify themselves permitted a finer analysis of the results (by comparison of subgroups) than would otherwise have been the case. Although the results were not especially clearcut, certain patterns could be discerned and recommendations formulated that had some basis in quantitative evidence.

7.10.4 Plenary discussion

In terms of information, relatively little came out of the plenary discussion following the administration of the questionnaire (see 7.7.3, above). Nevertheless, on psychological grounds there would seem to be some justification for offering a short period of up to 30 minutes during which participants can amplify their responses to items in a questionnaire or voice other concerns or comments. It is obviously desirable that tutors should not be present during this phase; as a logical extension of the participant-centred philosophy underlying this particular course and the design of the questionnaire, it was also desirable that the evaluator should be guided by the participants' agenda and employ questions only to establish the extent of agreement on particular issues or for purposes of clarification.

7.10.5 Informal comments

Very little input to evaluation came from the informal pre-session conversations (7.4.3). These may have been too brief or too public; it may be that post-session conversations (in a pub, for instance) would have proved more productive. Perhaps the important point is that these were opportunities for individuals to express strong feelings, if they had any, to a neutral party. Other people in other circumstances might have taken such an opportunity. On a positive note, the conversations did permit social interaction between myself and the group and to that extent may have reduced any feelings of distrust related to my role as evaluator.

7.10.6 The use of a non-participant observer

In conversations with PW, it had become clear that my involvement as evaluator was welcomed because hitherto, as PW put it, no one had had the time, interest or expertise to carry out a thorough evaluation of this (or any other FL INSET activity) and because, as

indicated in Appendix 7.1, PW's own post-course follow-up in the schools was necessarily spasmodic.

In many cases, it may be impossible for a course organiser to arrange for an observer to be present at all sessions and to coordinate the evaluation of the course. The experience represented by this case study suggests that if this is at all feasible it should be considered. Having an investment neither in the delivery nor the planning of the sessions, the observer is freer than either tutor or participants to monitor reactions, including his/her own, and can make judgements about observable rather than desired processes and effects. As points 9-14 and particularly 12-14 in 7.5.7.3 indicate, many of these judgements could simply not have been made by a course tutor teaching a session or merely by reference to participant journals. As this case illustrates, the combining of observer and evaluator functions can contribute directly to changes in a programme. (If data analysis and feedback to tutors take place while the course is in progress, this can also result in planning decisions being revised in respect of that course.)

The status (independent or otherwise), conduct and even the positioning of the observer during the sessions may, of course, have some effect on how s/he is viewed and therefore on the reliability and validity of the data collected. The care taken to establish my role and my physically detached non-involvement (see 7.4.2) were intended to send out clear signals and it appears that these were understood. Asked during the plenary discussion what their perceptions had been of me as an observer, participants agreed that I had seemed objective and non-intrusive.

7.10.7 Triangulation

One source of evaluative data is the observer's own experience of the course. However, by dint of informant triangulation, taking soundings from participants and tutors too, the observer-evaluator is in a position to ascertain whether perceptions are shared. The present study used method triangulation to take this a step further, by establishing through the questionnaire the *extent* to which perceptions of the course were shared. Since the journal data was rather limited, this procedure was not merely a second check but rather an attempt to confirm hypotheses. The result was a set of recommendations which appeared to

have considerable face validity and which were subsequently acted on, in large measure, as indicated in Section 7.9.

CHAPTER 8

CASE STUDY 4

A CLOSED GROUP COURSE FOR ESOL TEACHERS AND TRAINERS[1]

8.1 INTRODUCTION

Teacher trainers, like any other teachers, are change agents. That is, they are expected to bring about changes in those they teach: changes in terms of knowledge or skill, awareness or attitude. It may be possible to observe certain types of change during a course (and tests and examinations are designed to do just this). However, some of the changes that trainers might wish to see taking place in teachers relate to teaching practices - to things being done which were not done before the course or to things being done better - and although it may be possible to observe an increase in skill during a course (e.g. as a result of microteaching practice) there is no guarantee that this will carry over to the individual's own teaching context, that new ideas will be implemented or that there will be any long-term effect. Essentially, this is the argument for post-course evaluation.

One of the striking findings of the survey of UK institutions described in Chapter 4 was that there was very little evidence of systematic attempts to carry out post-course evaluation. The reasons for this seem fairly clear. On the one hand, there are the kinds of logistical difficulties referred to in that chapter and in Chapter 2, but on the other there may also be a certain lack of awareness of the importance, for summative purposes but also for formative purposes, of post-course evaluation. In Chapter 6, it was suggested that even where post-course evaluation is not strictly necessary it may still be informative. The

---

[1] A version of this chapter was published as 'Participant action plans and the evaluation of teachers' courses' in *Edinburgh Working Papers in Applied Linguistics*, Volume 7 (1996): 85-99

view taken here, then, is that the difficulties should not deter us from at least trying to carry out some form of post-course evaluation - and exploring methods of doing so.

Evidence from the survey (section 4.6.3.4) indicates that the institutions who attempt post-course evaluation tend to obtain their data in one or more of the following ways: through questionnaires sent to participants some months after the end of the course; from sponsors' reports on the participants or containing feedback from the participants; and from visits to the participants' home country and interviews, group discussion and observation there.

The literature contains few examples of post-course evaluation. Among these, Ward *et al.*'s (1995) account of their attempts to collect data on programme effects by means of interviews and videorecorded observation as well as questionnaires stands out for its thoroughness and rarity. A rather different approach is described by Alderson (1985a). Alderson's 'implementation plan', as described in Chapter 3, requires participants on a 10-week course to complete 'a personal plan of action for further ... curriculum development' (1985a: 148) to be followed up some months later. Unfortunately, Alderson's paper includes neither a documented account of results nor an indication of the difficulties involved in using this type of instrument for evaluation purposes.

The study reported in this chapter was a systematic attempt to gather information for programme evaluation purposes, based on the procedure described in Alderson (*op. cit.*) but using what will be referred to as participant action plans.

The research question which was the underlying focus of the study was as follows:

What are the pros and cons of participant action plans as a programme evaluation instrument ?

8.2 THE STUDY

8.2.1 Course participants

In October 1994, 15 teachers and teacher-trainers from Hamburg and Schleswig-Holstein attended a two-week tailor-made refresher course at IALS. The group consisted of six males and 9 females, all of whom were very experienced, the average age being about 40.

8.2.2 Background to the course

The course resulted from an approach to IALS through a third person who outlined basic information on the likely composition of the group and their wishes in relation to duration and intensity (a mornings-only programme) and course content (sessions on the language, cultural issues and the methodology of in-service training). These suggestions were subsequently refined in correspondence with the course leader, who approved the final timetable. It was agreed that though individuals would be carrying out self-directed project work in the afternoons, this would be prepared for and followed up in morning sessions.

8.2.3 Course content

The basic timetable consisted of 20 x 1½-hour sessions, taught in the mornings. There were workshops on language through literature, drama activities for ELT, teaching mixed levels and developing learning independence; three sessions on the methodology of in-service training; culturally-oriented lectures; workshops on language awareness (e.g. political correctness); and a session on self-evaluation. Afternoons were taken up by self-directed project work (prepared for and followed up, as agreed, in morning sessions), optional lectures, and a visit to Moray House Institute of Education in Edinburgh. Although I was the Course Director and handled 14 of the 20 morning slots, six other staff were involved, each contributing one lecture or workshop.

8.2.4 Evaluation questions

Most of the courses run at IALS form part of continuing programmes (open-access, certificated or sponsored) and evaluation therefore serves the purpose of programme development. In this case, the likelihood of continuation was uncertain; nevertheless, there seemed a value in evaluating the extent to which the course had appeared to meet the needs of the group and, in a broader sense, the use they made of it. This information could after

all be used in designing courses for any future group with similar characteristics. The evaluation questions were therefore as follows:

1. *Did the course meet participants' perceived needs ?*
2. *Did participants make subsequent use of what they had gained from the course ?*

8.2.5 Stages in the study

8.2.5.1 Writing the action plans

The idea of action plans was introduced to participants on the last day of the course. They were first taken through page 1 of the handout (Figure 8.1, below) and asked to complete page 2 (Figure 8.2), in class. The completed forms were then photocopied and the originals given back to the participants. They were told that in two or three months' time they would be asked to report on their progress towards the objectives they had formulated for the action plan.

There were no queries and no obvious resistance. Although some participants looked at their timetables or files, presumably as an aide-memoire, the plans were completed in 10-15 minutes.

ACTION PLANS

What is an action plan ?

It's basically a set of resolutions: what you will DO as a result of an experience, in this case a course. DOING may mean discussing with colleagues some of the ideas you've been exposed to; trying out something new in the classroom; reading; or something else. You decide what actions you want to take.

Why make an action plan ?

An action plan is a kind of bridge between theory and practice or idea and implementation. Before you can make an action plan, you will need to go through a process of review and evaluation. It goes without saying that this is valuable in itself, but if it then leads to concrete decisions about what you will DO, there is even more chance that you will apply something of what you have learned.

What are the criteria for an action plan ?

Good action plans distinguish clearly between short- and long-term goals. The goals are specified in concrete terms, and they are realisable.

Now turn over the page and try to formulate at least three concrete, realisable goals which derive in some way from your experience in Edinburgh.

Figure 8.1: Action plan, page 1: briefing notes

MY ACTION PLAN

I will definitely ...

1

2.

3.

I also hope to ...

4.

5.

6.

NAME .... DATE ....

ADDRESS ....

....

....

Figure 8.2: Action plan, page 2

8.2.5.2 Follow-up

A follow-up letter was sent out in December, some two months after the end of the course (Appendix 8.1). It combined Christmas/New Year greetings with an enquiry concerning progress with the points in the individual's action plan, a copy of which was enclosed with the letter. The letter crossed with a handful of seasonal greetings from Germany. No reference was made in any of these to the action plan.

In early February I received one letter responding directly to the issues raised in mine, and waited for the others.

By the third week in February I was feeling somewhat despondent and mentioned this to one of my colleagues. He suggested a number of reasons for participants' failure to reply (e.g. the course had been very short; some effects may only manifest themselves after a considerable period of time, if at all; participants might be embarrassed if they had nothing to report). The conversation prompted me to write another standard letter, at the end of February.

In the second letter (Appendix 8.2) I summarised the conversation with my colleague and included an extract from the one reply I had received. Slightly personalised versions of the standard letter were sent to the people who had written to me at Christmas. I also wrote back to the one participant who had replied to my letter.

My second attempt produced five replies within two to three weeks (four letters and one card, the latter promising a fuller response in due course - which never actually arrived). Only two of these were from people who had written to me at Christmas. One of my letters came back: 'return to sender, address unknown'. The number of the house turned out to have been wrong; the letter was sent out again, and a faxed reply was received almost immediately. By the end of March, three more responses had arrived and one more, written during the Easter holidays, at the end of April. This gave me a total of ten, two-thirds of the group.

Rejecting as too assertive the options of writing another letter or making a telephone call to those who had not responded, I decided on a picture postcard with a photo of Edinburgh on one side and "Hope to hear from you some time!" on the other. The five postcards went out in late June, just before the summer holidays. I received one response - also on a postcard - in early November, roughly a year after the end of the course and ten months after the first follow-up.

8.2.6 Data Analysis

The data thus fell into two categories: the 15 action plans, and the 11 responses to my letters and postcard concerning the action plans. With the exception of the two postcards,

responses were quite long, the majority occupying one typed page of A4, and the handwritten responses ranging from two to four pages in length.

The plans were scrutinised for specific reference to topics treated during the course; other points were also listed, grouped and categorised.

The letters and cards sent by participants in response to my letters were first analysed for the following:

- evidence that action had been taken on the action points formulated at the end of the course
- evidence of any other effects of the course.

Some respondents do refer directly and systematically to the points in their action plan; some refer to certain points but not others; and the minority write in such general terms that it is impossible to relate their comments directly to specific action points.

The quantified results are, then, incomplete in that self reports are available for only two-thirds of the group, and may even be incomplete in relation to those who have responded. In this sense, they are at best an approximation of reality or, to present the case more positively, an indication of minimum effects. The truth that they represent (and the reliability of this kind of evidence is discussed below, in 8.3.2.3) holds good for the time at which these soundings were taken; in due course, individuals may decide (or be able) to pursue ideas they had not hitherto explored.

8.2.7 Results

8.2.7.1 Data analysis stage 1: the action plans

Topic areas had been selected (and approved by the course leader) on the assumption that they would be of interest to a well-experienced group. Nevertheless, given the variety of experience and teaching contexts within the group it was unlikely that any one topic would appeal equally to everyone, and a range of course topics figured in the action plans, as indicated in Table 8.1, below. Items 1-6 in the table were named timetable sessions (these constitute, in fact, all the specific methodology sessions). Item 7 makes reference to a

special session on the final day which incorporated a demonstration of the Silent Way, and item 8 to activities on the first day and at the start of many other sessions.

Table 8.1: Content areas featured in course and figuring in action plans (n = 15)

1. developing learner independence	6
2. the use of literature in language teaching	5
3. drama techniques in language teaching	5
4. teaching mixed levels	4
5. project work	3
6. self-evaluation	2
7. new approaches (incl. work with Cuisenaire rods)	2
8. ice breakers	2

It is possible that more than the numbers indicated found these sessions interesting or of some value; what the action plans indicate, however, is that at the point the course ended these particular individuals felt sufficiently stimulated to want to do something about this interest. In this sense, although the action plan may be a useful indicator of the perceived relevance of course content, it goes beyond the usual kind of end-of-course questionnaire which asks whether participants find particular sessions interesting or valuable or which sessions they have found most valuable (*cf* Appendix 6.7).

Table 8.2, below, relates to more general or less predictable effects of the course.

Table 8.2: Action plan objectives related less directly to course content

1. change way of teaching/approach	4
2. use materials given	2
3. use materials collected	1
4. study books bought	1
5. share with colleagues	6
6. stay in contact with group	3
7. maintain connections with Edinburgh	2
8. pursue interest in dyslexia	2
9. pursue interest in comparing paintings and poems	1
10. teach folkdances to students	1
11. incorporate ideas in own publications	1
12. not continue to do every kind of non-teaching job	1

Objectives 1-4 differ from those in Table 8.1 in that, though related to pedagogical concerns, they are more general. 'Change way of teaching/approach' (item 1) includes both general statements such as 'be more creative' and very specific intentions, such as making use of a specific (but unspecified) technique or getting students to talk more. 'Use

materials collected' (3) refers to materials collected in the course of individual project work.

The second subgroup of objectives, 5-7, are interpersonal. 'Share with colleagues' (5) points to the possible spread of ideas beyond their immediate recipients, an important consideration in the case of participants who are teachers and teacher trainers) and therefore another effect that can only be evaluated some time after the course is over. 'Stay in contact with the group' (6) would seem to be one indicator of the cohesiveness of the group, which has in fact met on at least one occasion since the course.

The third subgroup, 8-12, relate to personal objectives only tangentially (if at all) related to course content. Project work (see Table 8.1) was an opportunity for participants to pursue existing interests (dyslexia (8) in one case; comparison of poems and paintings as teaching material (9) in another); other interests (folkdances, 10) or personal resolutions (11-12) appear to have been stimulated by the course itself.

8.2.7.2 Data analysis stage 2: comparison of action plans and participant reports on action taken on these

Analysis of participant reports on the action taken on their plans proved much more difficult than would have been the case if (a) each individual's starting point had been the same (b) they had been responding to a standard proforma. (See Appendix 8.3 for examples of action plans and reports corresponding to these.)

Given these two built-in variables, the size of the total sample and the lack of information provided in several of the responses, results are numerically speaking very inconclusive.

The results are presented in the same order as those for in Table 8.1.

Table 8.3: Number of participants taking action on specific objectives related to course content

Objective	Action	
	planned	taken
1. developing learner independence	6	3
2. the use of literature in language teaching	5	1
3. drama techniques in language teaching	5	1
4. teaching mixed levels	4	0
5. project work	3	0
6. self-evaluation	2	0
7. new approaches (incl. work with Cuisenaire rods)	2	1
8. ice breakers	2	2
	29	8

Table 8.4: Number of participants taking action on objectives less directly related to course content.

Objective	Action	
	planned	taken
1. change way of teaching/approach	4	3
2. use materials given	2	1
3. use materials collected	1	0
4. study books bought	1	0
5. share with colleagues	6	1
6. stay in contact with group	3	0
7. pursue interest in dyslexia	2	1
8. pursue interest in comparing paintings and poems	1	0
9. teach folkdances to students	1	0
10. incorporate ideas in own publications	1	0
11. not continue to do every kind of non-teaching job	1	1
	23	7

Based on participants' self-reports, there is then some quantitative evidence that action was taken in relation both to plans related directly and less directly to course content: the ratios for action taken to action planned - based on the totals in Tables 8.3 and 8.4 - being 1 : 3.6 and 1 : 3.3 respectively. If one also draws on qualitative data - what respondents actually say in their letters - there seems reason to suppose that these figures are an under-representation of action in progress/potential action. For instance, in relation to the teaching of mixed levels in Table 8.3 on which no action appears to have been taken, Respondent D writes: 'I find myself more prepared to care for the needs of the "differently gifted"'. Project work, being a general category, may have manifested itself in action in any one of a number of ways; there is certainly some overlap between this and the more personal objectives included in Table 8.4. Self-evaluation, similarly, is a rather general

category and, in a sense, underpins the whole enterprise. Several respondents comment specifically on this:

> 'I haven't done any thorough self-evaluation though it's in the back of my head and I seem to realize the mistakes I do over and over again and to look at them more objectively' (E)

> 'In my own courses I'm actually just beginning to work a lot more with different forms of evaluation' (F - a trainer)

> 'As I knew your letter would come - I hadn't expected it as soon as this though - I had already tried to watch myself a bit more closely in my teaching behaviour' (I).

In connection with the objectives expressed in Tables 8.4 and 8.5, there are comments on postponed meetings with colleagues; there are also several interesting references to changes in teaching approach (reflected in the emphases added to the quotations below):

> 'by now I hope I have understood that I can't become an entirely new teacher within a fortnight, that this is a slow process which will take a long time, which can only be done in small steps and will perhaps never be fully completed. **What I have already started to do is trying to provide phases in exercises which are less teacher-centred'** (I)

> 'Even before the course ... terms such as learner independence ... were the rage in this part of the world. I had heard of these and read about them, even used them in my classes. Owing to your course, however, it has become clear to me that they must play an even more dominant role than I had thought ... As a result ... **I have been constantly on the look-out for ways and means to bring about more learner independence** ...[goes on to illustrate this in relation to his language classes]
>
> While working with probationary teachers in my seminars, too, **I have given prominence to learner independence by making them draw on their considerable experience and having them discuss things amongst themselves before I feed in additional material and ideas.** We find this very satisfactory' (B - a trainer)

Similarly:

> 'Both target groups [school pupils and trainees] showed their spontaneous appreciation of my somewhat different teaching approaches after my return to school and to the seminar....All in all, those two weeks gave me a new impetus, which is still evident, I guess. I found out that **I tend to employ more student-oriented methods now. (Motto: "Let them discover things ... I'll help them" rather than "That's the way it is. Got the message?").** I had always thought that these methods would be rather time-consuming ... but they are not, if they are well dosed ... I think that's the secret of successful teaching - it's the mixture or better combination of the traditional (which wasn't that bad if you left out the extremes) and the new. Nothing is more harmful and demotivating than routine and patterns.' (C - a trainer)

8.2.7.3 Data analysis stage 3: explanations and other comments

At this stage of the analysis, two questions arose:

1. why had certain action points not been implemented: was this an individual matter (were some participants simply less active than others) or was it something to do with the nature of the action point, the way this had been formulated, or the level of priority assigned to it ?
2. why had only one person responded to the first letter ? To what extent were the hypotheses expressed in the second letter correct (bad timing; longer period necessary for implementation or for effects to become apparent; unrealistic expectations concerning participants' sense of responsibility)?

Non-implementation of action points

Analysis of self-reported individual action (Table 8.5, below) indicates that the majority of participants responding to the follow-up have achieved a proportion of the objectives they set for themselves. Where they have not, this appears to have been for reasons beyond their control (e.g postponement of meetings with colleagues); lack of opportunity (action points which relate to more distant future or syllabus areas do not lend themselves to use of specific ideas); changed awareness (realisation that planned actions were not appropriate); in some cases, personal problems also loom large.

In Table 8.5, below, participants responding by letter are referred to as LA, LB, etc; those who sent postcards as PJ and PK. Column 2 indicates the number of points on which action has reportedly been taken, figures in brackets being the number of points originally formulated by each individual. Action points not referred to in the response are noted under NO REF. (The full responses of LA, LB, LD and LE can be found in Appendix 8.4.)

Table 8.5: Individual implementation of action plans

	ACTION	NO REF.	NO ACTION	EXPLANATION/COMMENT
LA	4 (5)	1		
LB	1 (3)	2		
LC	3 (3)			but used material his way
LD	3 (6)	2		
LE	1 (6)		1	2 meetings cancelled; 'in my mind ...' to do certain things
LF	2 (4)			remaining objectives felt to be inappropriate
LG	0 (6)	6		personal problems; also new to training
LH	2 (4)	2		
LI	2 (6)	2	2	one ref. to future plan; reality of teaching context makes implementation problematic
PJ	(2)	2		
PK	(2)	2		

A further analysis, this time of the distribution of points across the two sections of the action plan (*I will definitely .../I also hope to ...*) showed no pattern either in relation to the inclusion of specific, short-term points under the first section and more general, long-term objectives under the second or in relation to action actually taken. As will be clear from Table 8.5, most participants expressed fewer than the 'maximum' of 6 objectives; two wrote nothing under 'I also hope to ...'.

Non-immediate response

Several respondents mentioned how busy they had been. Sometimes the pressures appear to be a combination of the personal and professional:

> 'three lively children, a sweet old granny, a house, a garden, a dainty cat and a job don't allow much time for self-evaluation' (H)

In other cases, the pressures of work are overwhelming:

> (school, seminar [duties in connection with training course], night class, private tuition, work for a publisher, and just started a computer course) 'when your letter arrived I was up to my neck in it... I find that my usual daily routine looks as follows: getting up, WORK, feeling terribly tired, going to bed' (D)

One person had actually started to reply but ' I still wanted to add and correct something. I was afraid you might find it too superficial'. She goes on: 'I even tried to phone you to stop you worrying ... Maybe we are just worried that we can't fulfil your expectations' (H).

Another points out that it takes time to assess the nature of the impact of a course:

> 'Maybe the reason for not responding earlier could be found in the fact that you need time to find out whether your teaching approaches, behaviour, etc have really changed' (C)

See also comments 3.1-3.5 in Appendix 8.4.

The reports also yield other information of both a more specific and a more general nature, among which are comments on the value of the action plan itself. This, it will be recalled, was referred to specifically in the first letter.

8.2.7.4 The action plan

Three respondents commented directly on the action plan itself and a fourth on its implicit effect in getting him to evaluate more than he might have otherwise.

> 'I really am grateful for the action plan you made us fill in. It definitely helped me to be more decisive in telling people what I'm willing and what I'm no longer willing to do. Which led to the fact that two of my colleagues took over jobs I'd been doing. I realise that I have again given in to the daily routine, though' (A).

She continues:

> 'My answer may be too late for your purposes, I'm afraid, but for me it turns out to be just in time. Evidently it is worth looking at one's action plans from time to time.'

The second writes:

> 'I liked the idea of the action plan because it forces you to make your ideas more obliging (?) ... [and] to ask again some weeks later if the planned action has already taken place or why it hasn't' (F)

And the third wonders whether it might not be a good idea to repeat the process a year on:

> As a short-term review, I daresay, my way of approaching subjects both with my students and trainee teachers has changed considerably ... Whether it will remain like this has to be seen and re-examined after, say, a year or so. (I'll gladly give you another report then.)' (C)

8.2.8 Conclusions

Given the mixed nature of the group, and particularly the presence of two subgroups (the teachers and the trainers), the course was designed to appeal not to all of the group all of the time, nor indeed to all of the group some of the time but to some of the group all of the time. In this, it seems to have succeeded. Table 8.1, above, indicates that, to judge by the action plans completed on the final day of the course, all eight content areas were valued by at least some members of the group. More significantly, action was subsequently taken in relation to five of these areas (Table 8.3); as noted in the comments following that table, these figures may be an under-representation of action in progress or potential action. More significant still may be the fact that the implementation that is reported goes beyond the mere adoption of teaching materials or ideas, extending to changes in understanding, attitude and approach (see, e.g. the quotations in 8.2.7.2, above.

Comments of the kind quoted in earlier sections have potential value for developmental purposes. They provide a window for the evaluator on the implementation process, access to insights that - together with information on constraints on implementation - can inform course-design in relation to aims, content and method or even serve as illustrative material for use on future courses. Their value for research purposes is taken up in section 8.3, below.

8.3 IMPLICATIONS OF THE CASE STUDY

8.3.1 The action plan as programme evaluation instrument

On the basis of the evidence gained from the first stage of analysis, action plans do seem to offer a feasible and economical way of gathering information that is of value to course organisers, especially as input to decisions concerning the content of a course for a similar group. It would be possible to draw the line at this stage, without further systematic follow up. This would, however, leave us as ignorant as we normally are about what effects - if any - are triggered in course participants. Moreover, participants would not feel the kind of positive pressure that some mention to go on evaluating themselves.

One of the strongest impressions left by reading participants' responses to the follow-up is that the pressures on teachers/trainers are such that although they may be able to slot certain new ideas into their everyday teaching (Appendix 8.4: 3.10), it is extremely difficult for them to find the time to adapt materials or to think through a different way of doing things. Despite this, there are signs that the process of adaptation has taken place or is taking place and that, in some cases, effects may prove quite long-lasting.

The qualitative data on which these conclusions are based has certain advantages over the kinds of quantitative data that might have been obtained by a standardised and closed questionnaire. Not only is it more informative, but it is also more reliable: the detail lends credence to the report.

These positive outcomes notwithstanding, the study has shown that to use this method of obtaining data requires a good deal of persistence. Even when data has been obtained, this is non-verifiable and may well be tantalisingly incomplete. Moreover, the process of data-analysis is time-consuming and the findings too various to be generalisable. There must also be doubts about the suitability of the method for large-scale use (i.e. on large populations or multiple courses). Seen strictly as an evaluation instrument, then, the action plan in its extended report-back form and as used in this study has considerable limitations. Does it therefore merit serious consideration ? The following section takes a critical look at some of the issues involved, practical and theoretical.

8.3.2 Issues

8.3.2.1 Negotiation of action plans

Within the framework described by Alderson (1985a), action plans are negotiated with participants, who are asked to predict any possible difficulties of implementation and consider how these might be overcome. The objective, as is obvious, is that the action points should be as feasible as possible, and the assumption is that this is more likely if there has been prior reflection and discussion. On short courses, when new ideas may be introduced as late as the penultimate day, it may be difficult to find an appropriate time for negotiation, especially since more than one meeting might be involved, although the kind of *reflection* that is essential could presumably be externally prompted if the action plan were introduced on the first day of the course. The suggestion made here is that although action plans are ideally the result of a period of negotiation, in this case the absence of negotiation was not necessarily a grave disadvantage. (It was hoped that the distinction between *'I will definitely ...' and 'I hope to ...'* would serve a purpose similar to that of negotiation in the Alderson model; in the event, this distinction seems to have been largely ignored.) It seems reasonable to expect that most participants naturally reflect on the relevance of the ideas presented to them, without being pressed to do so, and that since on short courses the range of possible action points (at least those directly related to the course) is restricted decisions will be easier.

One virtue of the action plan over the kinds of thinking that course participants do naturally is that it provides a framework and a focus for thinking about applications. As a statement of individual intent, however, it probably has little more significance than a set of New Year Resolutions, earnestly meant, but soon forgotten. What distinguishes it from the latter, with or without negotiation, is the promise of follow-up, which creates a form of mutual commitment.

Two modifications to the procedure adopted, both of which have a bearing on commitment, are worth considering. For instance, once individuals have formulated action points in draft form, these could be discussed with others in the group. This would not only reduce the likelihood of unrealistic objectives being put forward, it would also allow for the possibility of cooperative decisions, which might strengthen individual resolve and

make subsequent exchanges on progress between participants more likely. The proforma could also be redesigned. One possibility is shown below (a foolscap layout would be preferable to allow more space for the last two columns):

ACTION POINT	DATE	*Yes* (✓)	COMMENT	*No* - REASON?

Figure 8.3: Revised layout for action plan

Columns to the right of the action points would be completed by the participant at the time when the evaluator solicited a progress report. The new layout has three main advantages over the version used in the study. First, it includes space for a date - the date by which action on each point will be taken. This is completed at the time the plan is made. Responsibility for deciding the dates rests with the participant (the action plan *per se* may have been imposed by the evaluator, but decisions are taken by the individual concerned); and from a practical point of view, the dates alongside the action points serve as a clear guide to the evaluator concerning the timing of follow-up (see 8.3.2.2, below). Second, the 'Yes' column makes data analysis at the quantitative level a relatively simple matter. Third, the spaces for comment and explanation seek to elicit information that might be of value for programme development purposes. In the present study there are many gaps between the action plans and the reports. If we knew more about the conditions that encourage or militate against the fulfilment of action points (and this would obviously necessitate a a broader set of data-collection instruments), we would be in a better position not simply to help participants formulate realistic objectives but also design courses that are more approprioately targeted.

8.3.2.2 Timing of follow up

Difficulties relating to the timing of follow up were referred to in Chapter 2. In the case of action plans, there are perhaps two issues. One is the practical problem of selecting a time for follow-up which would not be obviously inappropriate (e.g. an exam or report-writing period) - the solution to which might be to ask participants for advice. The other is the length of time needed for gestation, for reflection to be translated into action, if action is what is expected (to which the date element in the format proposed above might be a partial solution). The difficulty here is that certain kinds of change may not take place for

several months or even years while others (practical ideas) might be implemented immediately. This implies the need for a longitudinal study continuing well beyond the period of the study described here. In such a case, it would obviously be desirable to combine individual contact by letter with other forms of contact (e.g. a newsletter, follow-up meeting with whole group, if this is possible). Apart from fulfilling an evaluative function, such contact would encourage the exchange of experiences and might well stimulate further, new action.

8.3.2.3 Reliability of self-reports

Where the desired outcomes of a programme can be specified in terms of behavioural effects, as might be the case in a *training* programme, and these are quantitative (something happens or it happens more frequently), observation is the most appropriate form of evaluation. However, the kinds of change in behaviour promoted by an *educational* programme are motivated by a change in perception or attitude. In such a case, although the possibility of misrepresentation, unwitting or otherwise, exists, a written self report may give some clue to the nature of the change and its effects (see e.g. the responses from B and C in 8.2.5.3 and the complete responses from LA and LB in Appendix 8.3).

The point has been made on a number of occasions in this thesis, however, that we do well to be a little cautious about accepting self-reports (in questionnaires, interviews or action plans) at face-value, especially when there is a possibility of loss of face for participant or tutor. In this particular case, although there is no reason to doubt that participants are telling the truth, and although nothing appears to hang on admissions that actions have not yet been taken, affective factors and role relationships may still play a part. Whereas some respondents cite factors beyond their control as a reason for late replies or non-action (Appendix 8.4: 3.2, 3.4), 'face' and concerns about the tutor-evaluator's expectations may be a factor in other cases (see, e.g. H's explanation (in 8.2.7.3) for her delayed response) and this may have been a consideration for those participants who did not make a return. A further step in such studies, as recommended in the literature surveyed in Chapter 2, might be to send such results as are available to participants in order to ensure that it represents as accurately as possible their motivations and interpretations as well as those of the evaluator. This would also be an appropriate point, obviously, to enquire what actions,

if any, have been taken since the last soundings were taken and thereby gain some insight into the time necessary for implementation to take place.

8.3.2.4 Gathering data on cause and effect

In assessing the relationship between effects and their assumed cause (the course), and therefore the validity of any conclusions based on the self-reports, it is necessary to consider the possible influence of other factors. Where the period of time between the course and subsequent implementation is a matter of months or even years, the certainty of a causal link is necessarily weaker. In this respect, however, self report - if it can be trusted - is potentially superior to observation. The observer is only in a position to see what happens (and perhaps whether it differs from what was happening previously); the ex-participant can ascribe a cause to that behaviour.

8.3.2.5 The managerial dimension

It will be clear from the foregoing discussion that action plans are not an easy answer to the problem of post-course evaluation. The resource implications of incorporating such plans (in the proposed form) into every course that is run would make a manager shudder. Time spent on course planning and implementation are patently necessary costs, and the amount of time required and when it will be used can in theory at least be predicted. Time will also normally be allocated for the meetings, data analysis and report-writing often associated with end-of-course evaluation. This is, in resource allocation planning terms, part of the total time allocation for the course. Post-course evaluation does not fit into this neat scheme. Rather than being perceived as an integral part of course design, it encroaches on the time that should be used for the next course. This might be less of a problem if the time required could be easily quantified and restricted to a certain period on the calendar. Where these calculations are impossible because of the individualised nature of the data collection method, as is the case with the approach to action plans taken in this study, managerial approval is highly unlikely.

One possibility would be to attempt to argue the case for experimentation in respect of a small-scale but longstanding programme which has not been reviewed for some time. Where the population for this programme is relatively stable, the results may be generalisable to other cohorts. An alternative would be to present action plans not as an evaluation measure but as a feature of programme provision which yields evaluation data

that would otherwise not have been available. The latter suggestion is developed in the next section.

8.3.3 The action plan as pedagogic device

Participants' reactions to the action plans suggest a second result of the study. With hindsight it is possible to argue that apart from any value that they might have for programme evaluation purposes, action plans serve a potentially crucial pedagogic purpose, by extending reflection on specific aspects of the course and their own teaching into participants' workplaces. In this respect, they serve a function similar to that of the Adviser (Ch.7) who follows up after an INSET course or within-course reflection prompted by the kinds of action plan described by Estaire (1993) or by journal-writing (Hundleby and Breet 1988; Green 1993).

To recognise this, however, is also to accept that the reports produced some time after the course are not simply reports on the effects of the course as experienced at a certain point in time, but testimony to the combined effects of the course and the action plan. It may even be the case that the course is perceived differently as a result of thinking about the action plan. At the level of the individual, the focus of evaluation has shifted. What is evaluated is no longer the course but the extent to which or the ways in which the individual makes use of the course experience. Course-evaluation is replaced by self-evaluation. This way of viewing the relationship between action plan and course is represented in Figure 8.4, below.

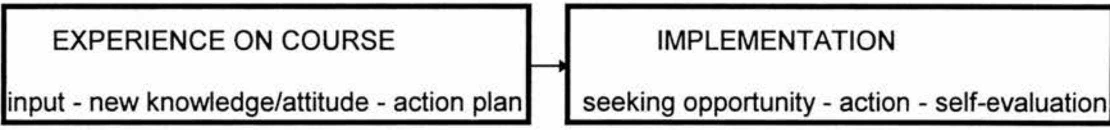


Figure 8.4: The action plan as stimulus to action and self-evaluation

8.3.4 Conclusion

On balance, the gains involved in the use of action plans in small-scale studies would seem to be worth the effort involved. In concrete terms the gains for the evaluator and institution take the form of increased awareness of what participants take from a course and the obstacles that stand in the way of implementation, and to some extent this information may

be generalisable. In asking - as we normally do - about the success or otherwise of a course as a whole, we may be asking the wrong question; we perhaps ought to be looking more carefully at the effect on individuals in relation to their personal/professional agendas (as well as that of the course provider, and in the case of sponsored groups, that of the sponsor). There is, in addition, the potential value to the participant of continuing contact with the providing institution. The inclusion of action plans in a programme, if taken seriously by the institution and followed up, is a recognition of shared responsibility with the participant for positive outcomes. Follow up may even prompt action that would not otherwise have taken place. Referring to post-course questionnaires, Alderson (1985a) notes that they are a sign of 'caring'. This dimension of caring on the one hand and the very individual nature of the information received from participants means that the action plan is probably as close as we can get to participant sensitive INSET and participant sensitive programme evaluation.

CHAPTER 9

SUMMARY AND CONCLUSIONS

9.1 INTRODUCTION

Compared to experimental, quantitatively oriented studies, the thesis has a rather 'soft-edged' and often tentative feel. There are probably three reasons for this. Firstly, the limited number of programme evaluation studies in the L2 teacher education literature and the descriptive emphasis of most of these studies made it difficult to relate my findings to those of others. Secondly, the limitations imposed on the case studies by the research questions I chose to ask (e.g. excluding observation from Case Study 1) or the incompleteness of the data (in Case Studies 2 and 4) meant that the findings were necessarily somewhat inconclusive. And thirdly, I selected research questions that would enable me to explore particular methods of investigation (non-participant observation, particiopant journals, action plans, document analysis) in order to develop my own competence as a researcher and an evaluator. With more extensive experience, I would perhaps feel more confident of making stronger claims. While just as sympathetic to naturalistic enquiry methods as I was at the outset, I am now more conscious of the difficulties involved in the collection, processing and analysis of data and in reporting. My learning is reflected in the recognition of, for instance, the need for a more structured approach to the use of journals and action plans, the fact that effects may take longer to appear than one expects, and the importance of careful record-keeping.

A substantial section of the thesis is taken up with the case studies. In the next section I discuss the *pros* and *cons* of case study research and relate these to the studies carried out as part of my own research. Section 9.3 then returns to the research questions posed in Chapter 1 and summarises the answers to these questions. Section 9.4 offers a number of general conclusions.

9.2 CASE STUDY RESEARCH

9.2.1 Advantages and disadvantages of case studies

Some of the arguments in favour of case studies are:

1. they are readily understandable (concrete rather than abstract, and usually written in a non-academic style);
2. they have a 'three-dimensional reality, like a good documentary' (Nisbet and Watt *op.cit.*: 76); they can reflect different points of view; and they lend themselves to the same kinds of judgement that are made in everyday life (Adelman, Jenkins and Kemmis 1994);
3. they provide a basis for comparison with (and interpretation of) other similar cases (Nisbet and Watt, *op.cit.*) or for generalisations about the case in point (Adelman et *al.*, *op cit.*);
4. the resulting data may be a useful resource for subsequent researchers with other purposes (Adelman *et al., op.cit.*);
5. they can contribute to action: for use in staff development, for educational policy-making and for formative evaluation (Adelman *et al., op.cit.*);
6. they may identify 'a pattern of influences too infrequent to be discernible by more traditional statistical analyses' (Nisbet and Watt *ibid.*);
7. they can be undertaken by a lone researcher (Nisbet and Watt, *op.cit.*).

On the other hand, as Adelman *et al.* (*op.cit.*: 96) have pointed out, 'Case studies are not easy to do. Some of our best friends are presently trapped inside case studies, trying to get out. Almost none will escape unscathed'. I am now in a position to vouch for the truth of these observations.

Their scale and complexity apart, the main disadvantages of case studies referred to in the literature are that:

1. considerable subjectivity is involved in the selection of what is reported and how it is reported;
2. their findings are not generalisable (which is in direct contrast with point 3, above).

As we have seen in earlier chapters, subjectivity in the form of selection is an element, conscious or otherwise, in all evaluation and at all stages of evaluation from planning to reporting. In case study, selectivity - a major problem for the evaluator - is clearly essential, but provided the data and the means by which it was obtained are open to scrutiny and there is evidence that the final account has been validated by those portrayed this should not be a problem for anyone else.

The main limitation of case studies from a methodological perspective seems to be the difficulty of generalising from a single instance; in other words, case studies would appear to lack any wider validity. Adelman *et al.* (*op.cit.*) offer an interesting response to this charge:

> Experimental research "guarantees" the veracity of its generalizations by reference to formal theories and hands them on intact to the reader; case study research offers a surrogate experience and invites the reader to underwrite the account, by appealing to his tacit knowledge of human situations. The truths contained in a successful case study report, like those in literature, are "guaranteed" by "the shock of recognition"
>
> (Adelman *et al.* 1984: 96)

(This assumes, of course, that researchers never lie and that the reader can identify with the case.)

Moreover, the assumption that research should necessarily lead to generalisations which are applicable to other cases may itself be suspect:

> It is tempting to argue that the accumulation of case studies allows theory building via tentative hypotheses culled from single instances. But the generalizations produced in case studies are no less legitimate when about the instance rather than about the class from which the instance is drawn (i.e. generalizing about the case, rather than from it).
>
> (Adelman *et al. op.cit.*: 94)

Bassey (1984), writing about pedagogic research as a subset of educational research, takes a stronger line, contesting the basic assumption that such research can or indeed should lead to useful generalisations. Bassey's argument rests on a distinction between 'closed' generalisations, which are decriptive and refer to a specific set of events, and 'open' generalisations, which are both descriptive and predictive (*op.cit.*: 111). The latter, he

claims, are so rare as to be an inappropriate object of pedagogic research; what interests the individual teacher is rather the 'relatability' of the 'closed' generalisations resulting from the research to the specific institution and classroom within which he finds himself., Case studies, which permit teachers to make their own judgements as to what is relevant and relatable, may be more valuable for this purpose, he argues, than research directed towards open generalisations since the latter inevitably omits the fine textural detail with which teachers may be able to identify.

Developing this idea, Adelman *et al*. point out that case study research can be set up in one of two ways: (1) the researcher works from a hypothesis or an issue and chooses an instance or case thought to be representative (2) the case forms the starting point, and through careful exploration the investigator identifies certain issues or influential factors.

Approach 1 tends to result in generalisations about the *class* from which the instance is drawn (although, as Adelman *et al*. point out, thinking presumably of the thickness of description, 'the description of the case will increasingly emphasize its uniqueness' (*op.cit.*: 95)). Approach 2 lends itself to generalisations about the *case*, and may permit generalisation to other cases. In practice, however, as in approach 1, the context-embeddedness of cases makes such generalisation difficult if not impossible.

One way of distinguishing between the two approaches is to see them as related to purpose or audience. Where the researcher's interest is in an issue, he is likely to use approach 1, whereas if his interest lies in a 'bounded system' (Adelman *et al., ibid.*), such as a specific course within a specific institution, he will be more likely to use approach 2. As I indicate at the end of this section, I made use of both approaches.

Reynolds (1982), who sees case study research as potentially generalisable, specifies two conditions:

> The prime requirements upon case study research are that the description be faithful to the situation described (i.e. valid) and that both the data and the analytic procedures are made public, so that the reliability of the analysis can be tested. If these criteria are met then the research account is available for comparison with accounts of other situations, application to similar situations and even incorporation into existing or developing theories. In this sense, case study research is generalizable: potentially if not actually so, by being publicly on offer.
>
> (Reynolds 1982: 55)

To some extent the argument rests on the question of the audience for the final report. The general context for the discussion in this thesis has been that of institution-based evaluation and the primary audience for a case study report would be the individual(s) responsible for programmes and programme personnel. In this case, the extent to which findings are generalisable to other programmes offered by the institution will obviously be what is of interest rather than its wider generalisability. However, it is also possible to identify a secondary audience. The suggestion made in Chapter 1 was that programme evaluation can have an educational purpose, by contributing to understanding within the profession at large, and with this audience in mind generalisability is seen to be a more significant issue.

Three of the case studies which form a substantial part of this thesis illustrate Adelman *et al.*'s Approach 1. The cases studied in Chapters 5, 6 and 8 were chosen as representative of particular classes of programme (in-house seminar as part of a TD programme, refresher course, tailormade course) and because they permitted the exploration of particular methods for use in programme evaluation. As a result, it is possible, as Adelman *et al.* suggest, to generalise about the class from which these cases are drawn *and the methods that can be used to evaluate such programmes*. Approach 2 is exemplified by the study of the Lothian Region course (Chapter 7), in which issues and influential factors only emerged over time. In principle, the context-embeddedness of such cases should make generalisation difficult if not impossible, as noted above. In practice, because the research focus of this case study, as of the others, was not the course but evaluation method, certain generalisations are possible, as signalled in the conclusion to that study.

9.2.2 Sensitive evaluation

A theme that runs through the discussion of each of the case studies is the importance of what might be termed 'sensitive evaluation'. The case study approach, which typically makes use of more than one method of data collection and more than one data source, is well suited to the discovery if not of absolute truth at least of overlaps of reality in relation to specific cases.

Case studies after all start from the assumption that they are dealing with a unique entity and the programmes run by an institution are unique at both a general and a more specific level. At the level of programme rather than course there is clearly a need for a thorough initial evaluation to check that, for example, the assumptions underlying the programme design are soundly based and administrative arrangements working properly (Case Study 3 - Lothian Region course). Beyond this, there is a need for regular review of the extent to which, for instance, the target population is stable, the objectives appropriate, the materials perceived as useful, and the programme evaluation procedures themselves still relevant (Case Study 2 - TLE). Within a programme there are courses. Since the dynamism of a course comes from those involved, tutors and participants, we can expect that each course will differ to a greater or lesser extent from that which preceded it, which implies the need for course sensitive evaluation, a need felt particularly by tutors (Case Study 2). Where a programme is a more loosely organised entity, consisting of an indeterminate progression of independent components, resources and common sense argue against the evaluation of each component; useful lessons might nevertheless be learned from the evaluation of component tokens (Case Study 1 - seminar on the OHP). Finally, there is a need for evaluation that is not just sensitive to individual participants (pre-/post-evaluation in Case Study 1; journals in Case Study 3) but works with participants' agendas (questionnaire in Case Study 3) and goes temporally beyond the normal end-of-course questionnaire to track individuals in their own contexts (Case Study 4 - course for German teachers and trainers). In relation to INSET, the last of these suggestions has particular importance. As Murphy (1995: 24) has pointed out, 'the solution is not a new approach [in teaching or teacher education] but a better understanding of the context where language teaching goes on'. Murphy is here thinking of the classroom; but evaluative information gathering needs to spread its net wider, to look at the factors within individuals, schools and societies which facilitate or militate against implementation. A study of general UK INSET by Evans and

Hopkins (1988, cited in Hopkins 1989) found that implementation was dependent on two factors: the nature of the individual teacher and the ethos within the school. Similar studies are needed of L2 INSET. In order to have an informed basis for future action we must make a greater effort to understand not only the impact of INSET programmes on teachers in their own contexts but also the impact that those teachers have on the institutions in which they work.

9.3 THE QUESTIONS REVISITED

In Chapter 1 a number of specific questions were posed. These were:

1. *What methods and techniques are available for the evaluation of INSET programmes for language teachers ?*

The literature review in Chapters 2 and 3 revealed that programme evaluation is typically based on data collected by three methods (questionnaire, interview, observation - each of which can take a variety of forms), and on the analysis of programme documents. Two further data-collection methods (participant journals, participant plans) and a small number of techniques that do not fall easily into the broad methods categories (e.g. posters, concept-mapping) were also identified.

2. *What is the situation within UK institutions with regard to the evaluation of INSET programmes for language teachers ? How are such programmes evaluated? Are there any patterns in the procedures used or not used ? How does actual practice compare with that recommended in the evaluation literature ? How aware are those responsible for evaluation of any inadequacies in approaches to evaluation within their institution ?*

The stated focus of the thesis was the evaluation of INSET programmes within UK institutions. The survey reported in Chapter 4 found a broad similarity between the data collection methods referred to in the literature and those most commonly used by institutions (with the questionnaire being the dominant method), although there was some variation in the extent of use according to programme-type. Individual interviews (tutorials) were used, but plenary discussion featured more frequently. Observation was

relatively underused within certain types of programme and as a measure of post-course effects. This corresponds to the finding that little post-course evaluation takes place, the major cause of concern to those survey respondents (located mainly within universities) to express any disquiet.

A further gap in evaluation systems - and one of which survey respondents seemed unaware - lies in the area of pre-/post- comparison. Either no effort is made to collect baseline data or, if such data is available, it is not used to assess or explain the degree of individual change.

3. *Of the procedures that do not appear to be commonly used, which might make a positive difference to the effectiveness or efficiency of evaluation ?*

The relatively closed nature of the postal questionnaire used in the survey limited its value as a source of new ideas and, indeed, as a source of information on how institutions operationalise particular data-collection methods. In this respect, the interviews (Appendix 4.6) afford a much clearer insight into what happens and how this differs from what appears to be the norm. Three specific procedures which do not appear to be widely used were noted as a result of the interviews:

1. diaries (i.e. participant journals), as a useful complement to other forms of data-gathering;
2. delayed post-course questionnaires, as one means of collecting data on programme impacts on participants;
3. the involvement of someone who is not a member of the course team.

In relation to (3), it should be noted that although the insider/outsider debate is frequently rehearsed in the literature, it is not, to judge by the questionnaire returns, a feature of the evaluation systems operated by most institutions.

These three procedures were then examined from particular perspectives in the Case Studies. Case Study 3 made use of participant journals and a non-participant observer as evaluator; Case Study 4 considered the potential of participant action plans, as an alternative to the questionnaire for post-course evaluation. Table 9.1 indicates the results obtained by these and other methods.

Table 9.1: The case studies: methods used and data collected

	questionnaire	interview	journal	action plan	observation	document analysis
CS 1	collection of baseline data	schedule used to obtain feedback on effects check for further effects				
CS 2	formulation of language-learning needs session ratings (RAF) feedback on P impressions of first week (letter) P feedback on various aspects of course	check match between course and P expectations and needs (tutorials) explanation/ amplification of questionnaire results	indication (limited) of positive/ negative reactions and concerns			insight (partial) into measures used over time and their relative contribution to formative and summative evaluation
CS 3	P feedback on various aspects of course	collection of background data and clarification of programme aims explanation/ amplification of questionnaire results	indication (limited) of positive/negative reactions and concerns		account of process (and judgements) which could be compared with those of others	course aims and content attendance figures as indicator of motivation level
CS 4				feedback on perceived applicability of course content subsequent reports provided feedback on implementation (and other effects) and constraints on implementation		

Chapter 1 also posed a more general question: 'What might properly organised evaluation procedures look like ?' Although there are obviously constraints on what is feasible, the argument that has been advanced here is that where relevant systematic data-gathering needs to start from a point before the course begins (to establish the baseline) and make use of that information retrospectively in relation to individuals as well as course cohorts (pre-/post-programme analysis); it also needs to continue (on a selective basis) beyond the end of the course and probably involve spaced checking over an extended period. Records are needed, of the sort that facilitate regular programme review, and these should include a note of action taken on evaluation reports. Evaluation procedures should themselves be included in these reviews. In establishing such a system for institutional programme evaluation, managers should be mindful of the decision-making model presented in Chapter 2.

9.4 CONCLUSIONS

Language teacher educators think of themselves, rightly, as professionals, but when it comes to programme evaluation, to judge from the survey reported in Chapter 4 and the conversations reported in Appendix 7.1, they start to feel a little uneasy. Some of the reasons for this have their root in the constraints from which most institutions suffer and do not necessarily reflect on the individuals. Such constraints do not, however, absolve teacher educators individually and in institutional groupings from the need to formulate a principled approach to the evaluation of the programmes in which they are involved and be as explicit about this as they are about any other aspect of their work. This will probably entail reading and even further training; ultimately, however, the most effective learning will result from *doing* evaluation and reflecting on the process and products.

Scriven (1996: 404) has drawn attention to the enormous significance of evaluation:

> Evaluation is not only a discipline on which all others depend [i.e it is what he calls a 'transdiscipline'], it is one on which all deliberate activity depends. It follows that significant improvements in the core concept and techniques of evaluation, of which we have seen many in recent years, and of which many more could be made within the next few years, have the potential for huge improvement in the quality of life and work, as well as in the level of achievement in all disciplines.

For these anticipated effects to be felt within UK institutions, two forms of action are needed. First, managers of institutions need to take steps to ensure that programme evaluation processes provide in an efficient manner the information necessary for purposes of accountability and development and that the information is used. In practice, this responsibility might be delegated to someone with special expertise or interest in evaluation (i.e. an Evaluation Officer), who would be charged with keeping up to date with developments in the field and advising other members of staff. Second, more evaluation studies are needed with which those involved in second language teacher education can identify. The case studies included in this thesis were conceived as a means of exploring the viability and benefits of particular evaluation methods rather than as evaluation studies in their own right (although Case Study 3 also served the latter purpose). Nevertheless, as a means by which outsiders can gain access to the inner workings of a programme and the forces (including evaluation) that shape it, the case study has much to commend it. Follow-up evaluation studies have already been mentioned as a focus that would have particular value for pedagogic purposes (i.e. programme development); meta-evaluation studies - as the thesis has argued - are also necessary. Increased awareness of what has been done and therefore can be done would be the first stage in a virtuous circle leading to improvement in the quality of programmes, evaluation systems and professional competence.

NOTES

Chapter 1

[1] The argument for post-course evaluation is as strong for PRESET as it is for INSET. Gaies (1992) describes preliminary steps towards a form of portfolio assessment which will include reports from programme participants after they graduate (presumably as part of a probationary requirenment), and Dellar (1990) and Davis (1991) show the potential for programme development of studies into the problems experienced by novice teachers.

[2] A useful summary of the history of UK INSET from the eighteenth century up to the mid-70s, seen alongside developments in the USA and Europe, is contained in Henderson (1978); accounts of more recent developments in England and Wales can be found in Mortimore and Mortimore (1989) and Gilroy and Day (1993).

[3] In fact, as Fox (1980) insightfully points out, the work carried out on curriculum evaluation by such evaluators as Parlett and Hamilton, Elliott, MacDonald , Walker, Stake and Scriven was also of relevance for those concerned with teacher education since it was centrally concerned with the issue of how new curricula are integrated into school practice.

[4] Mackay (1981) reports that in the course of a painstaking search in 1975 for guidance on the evaluation of ESP programmes he was able to locate only three relevant references, none of which provided a suitable evaluation model.

[5] Walker recalls that Parlett and Hamilton's (1972) paper on illuminative evaluation, 'having failed to find a journal publisher, was run off in an edition of some 200 copies, yet within months had circulated underground in numerous forms in what must have been at least a tenfold multiple of the original print run' (Walker 1985: 1-2).

[6] Despite its publication date, this volume was conceived in 1986-87 (Alderson, personal communication).

[7] A British Association of Evaluators was formed in November 1994.

[8] See, however, Woodward (1988), which was published by Pilgrims.

[9] The chapter on evaluation in Woodward (1991) was included at the suggestion of the series editor, Roger Bowers (Woodward, personal communication).

Chapter 2

[10] Scriven (1996) draws attention to other types of cost, such as 'evaluation stress'.

Chapter 3

[11] Denzin (1970) also distinguished time, space. theoretical and combined forms of triangulation (Cohen and Manion 1994).

Chapter 4

[12] The CD report highlights issues and proposed action. This is formulated into an action plan by the institution's Academic Standards Unit (ASU) and resulting action is reported by the CD in his/her next report and monitored by ASU. The procedure forms part of an institution-wide policy.

[13] Wallace (1997) and Kennedy (1997) both endorse the view of Handy (1995) who argues that most human enterprises follow a predictable cycle, with decline as a final stage, and that if we are to avoid this we need to set in motion a new cycle *before decline sets in*.

REFERENCES

Abbey, S. 1991. *A Case for On-going Evaluation in English Language Teaching Projects*. CLCS Occasional Paper No. 29. Dublin: Trinity College, Dublin

Adelman, G., Jenkins, D. and Kemmis, S. 1984. Rethinking case study. In Bell, J. *et al.* (eds) 1984: 93-102

Ainscough, V. 1994. Sustainable methodologies: a call for more sensitivity to relevance in ELT projects. *The Teacher Trainer* 8.2: 16-19

Alderson, J. 1985a. Is there life after the course ? In Alderson, J. (ed.) 1985b: 129-150

Alderson, J. (ed.) 1985b. *Evaluation*. Lancaster Practical Papers in English Language Education 6. Oxford: Pergamon

Alderson, J. 1992. Guidelines for the evaluation of language education. In Alderson, J. and Beretta, A. (eds) 1992.: 274-304

Alderson, J. and Beretta, A. (eds) 1992. *Evaluating Second Language Education*. Cambridge: CUP

Alderson, J. and Scott, M. 1992. Insiders, outsiders and participatory evaluation. In Alderson, J. and Beretta, A. (eds) 1992: 25-60

Alderson, J. and Waters, A. 1982. A course in testing and evaluation for ESP teachers or, "How bad were my tests ?" In Waters, A. (ed.) 1982. *Issues in ESP*. Lancaster Practical Papers in English Language Education 5. Oxford: Pergamon: 39-61

Alexander, R. 1980. The evaluation of advanced inservice courses for teachers: the challenge for providers. *British Journal of Teacher Education* 6.3: 177-195

Alkin, M. 1970. Evaluating the cost effectiveness of instructional programs. In Wittrock, M. and Wiley, D. (eds) 1970. *The Evaluation of Instruction: Issues and Problems*. N.Y.: Holt, Rinehart and Winston: 221-238

Alkin, M. 1991. Evaluation theory development. In McLaughlin and Phillips (eds) 1991: 91-112

Anivan, S. (ed.) 1991. *Issues in Language Programme Evaluation in the 1990's*. Anthology Series 27. Singapore: Regional Language Centre

Argyris, C. and Schön, D. 1974. *Theory in Practice: Increasing Professional Effectiveness*. San Francisco: Jossey Bass

Ashworth, M. 1985. *Beyond Methodology*. Cambridge: CUP

Aubrey, C. 1988. Guidelines for an effective evaluation. *British Journal of In-Service Education* 14.3: 140-146

Bachman, L. 1981. Formative evaluation in specific purpose program development. In Mackay, R. and Palmer, J. (eds) 1981: 106-16

Bailey, K. 1983. Competitiveness and anxiety in adult second language learning: looking at and through the diary studies. In Seliger, H. and Long, M. (eds) 1983. *Classroom Oriented Research in Second Language Acquisition*. Rowley, Mass.: Newbury House: 67-102

Bailey, K. 1990. The use of diary studies in teacher education programs. In Richards, J. and Nunan, D. (eds) 1990: 215-226

Bailey, K. and Ochsner, R. 1983. A methodological review of the diary studies: windmill tilting or social science ? In Bailey, K., Long, M. and Peck, S. (eds) 1983. *Second Language Acquisition Studies*. Rowley, Mass.: Newbury House: 188-197

Baker, K. 1980. Evaluation of the Teacher Induction Pilot Schemes (TIPS) Project: some problems of a largescale action research project. In McCabe, C. (ed.) 1980a: 106-115

Bartlett, L. 1990. Teacher development through reflective teaching. In Richards, J. and Nunan, D. (eds) 1990: 202-214

Bassey, M. 1984. Pedagogic research: on the relative merits of search for generalization and study of single events. In Bell, J. *et al.* (eds) 1984: 103-22

Bell, J. 1993 (2nd ed.) *Doing Your Research Project*. Buckingham: Open University Press

Bell, J., Bush, T,., Fox, A., Goodey, J. and Goulding, S. (eds) 1984. *Conducting Small-Scale Investigations in Educational Management*. London: Paul Chapman/Open University

Berdie, D. and Anderson, J. 1974. *Questionnaires: Design and Use*. Metuchen, New Jersey: The Scarecrow Press

Beretta, A. 1986a. A case for field-experimentation in program evaluation. *Language Learning* 36.3: 295-309

Beretta, A. 1986b. Towards a methodology of ESL program evaluation. *TESOL Quarterly* 20.1: 144-155

Beretta, A. 1986c. Program-fair language teaching evaluation. *TESOL Quarterly* 20.3: 431-444

Beretta, A. 1990a. Who should evaluate L2 programmes ? In Brumfit, C. and Mitchell, R. (eds) 1990. *Research in the Language Classroom*. ELT Documents 133. London: Modern English Publications/British Council: 155-160

Beretta, A. 1990b. The program evaluator: the ESL researcher without portfolio. *Applied Linguistics* 11.1: 1-15

Beretta, A. 1992. Evaluation of language education: an overview. In Alderson, J. and Beretta, A. (eds) 1992: 5-24

Beretta, A. and Davies, A. 1985. Evaluation of the Bangalore Project. *ELT Journal* 39.2: 121-7

Bloom, B., Engelhart, M., Furst, E., Hill, W. and Krathwohl, D. 1956. *Taxonomy of Educational Objectives*. Handbook 1: the Cognitive Domain. N.Y.: David McKay

Bolam, R. 1980. A national perspective. In McCabe, C. (ed.) 1980a: 39-49

Borg, W. and Gall, M. 1989 (5th ed.) *Educational Research*. N.Y.: Longman

Breen, M., Candlin, C., Dam. L. and Gabrielsen, G. 1989. The evolution of a teacher training programme. In Johnson, R. (ed.) 1989: 111-135

Brinton, D., Holten, C., and Goodwin, J. 1993. Responding to dialogue journals in teacher preparation: what's effective ? *TESOL Journal* 2.4: 15-19

British Association for Applied Linguistics. 1994. *Draft Recommendations on Good Practice in Applied Linguistics*

British Council. 1981. *Design, Evaluation and Testing in ELT Projects*. Dunford House Seminar Report 1980. London: The British Council

British Council. 1995. *Evaluation in Planning and Managing Language Education Projects*. PRODESS Colloquium, March 26-28, 1995, Manchester. Manchester: The British Council

Broady, E. 1995. 'You are your own best resource': promoting confidence and autonomous learning in teacher education at a distance - a case study. In Howard, R. and McGrath, I. (eds) 1995: 49-65

Brumfit, C. 1984. The Bangalore procedural syllabus. *ELT Journal* 38.4: 233-41

Buckley, R. and Caple, S. 1990. *The Theory and Practice of Training*. London: Kogan Page

Busch, M. 1993. Using Likert scales in L2 research. A researcher comments ... *TESOL Quarterly* 27.4: 733-736

Calder, J. 1994. *Programme Evaluation and Quality*. London: Kogan Page

Calderhead, J. 1990. Conceptualising and evaluating teachers' professional learning. *European Journal of Teacher Education* 13.3: 153-60

Chambers, F. and Erith, P. 1990. On justifying and evaluating aid-based ELT. *ELT Journal* 44.2: 138-43

Cohen, L. and Manion, L. 1994 (4th ed.) *Research Methods in Education*. London: Routledge

Coleman, H. 1992. Moving the goalposts: project evaluation in practice. In Alderson, J. and Beretta, A. (eds) 1992: 222-245

Cooper, J. 1983. Basic elements in teacher education program evaluation. In Howey, K. and Gardner, W. (eds) 1983. *The Evaluation of Teachers: a Look Ahead*. N.Y.: Longman: 118-135

Cronbach, L. 1963. Course improvement through evaluation. *Teachers College Record* 64: 672-83. Reprinted, with editorial changes, in Madaus et al (eds) 1983b: 101-115

Cronbach, L. 1975. Beyond the two disciplines of scientific psychology. *American Psychologist* 30.2: 110-127

Cronbach, L. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass

Cumaranatunge, L. 1989. Evaluating short term ELT teacher training courses: an HIEE experience. *F.O.E.* 5.2: 8-20

Cumming, A. 1988. Educating through educational evaluation: an idealistic response to Ronald Mackay's 'Programme evaluation and quality control'. *TESL Canada Journal* 5.2: 43-50

Davies, A. 1992. Review of Alderson, J. and Beretta, A. (eds) 'Evaluating Second Language Education'. *Language Testing* 9.2: 207-9

Davis, H. 1990. *The Four-Week CTEFLA Course: a First Step Towards Evaluation*. Cambridge: University of Cambridge Local Examinations Syndicate

Davis, J. 1980. The evaluation of Teachers' Centres and school-based activities. In McCabe, C. (ed.) 1980a: 94-105

Dean, J. and Whyte, W. 1958. How do you know if the informant is telling the truth ? *Human Organisation* 17.2: 34-38

Delbecq, A. and VandeVen, A. 1971. A group process model for problem identification and program planning. *Journal of Applied Behavioral Science* 7.4: 466-491

Dellar, G. 1990. The needs of 'novice' teachers: a case study. In Roberts, J. (ed.) 1990. *Initial Training and the First Year in School*. CALS Workpapers, Vol.1. Reading: Centre for Applied Linguistcis, University of Reading: 62-77

Dingwall, S. and Mann, S. (eds) 1982. *Methods and Problems in Doing Applied Linguistic Research*. Department of Linguistics and Modern English Language, University of Lancaster

Dock, A., Duncan, W. and Kotalawala, E. 1988. *Teaching Teachers through Distance Methods*. Colombo: Swedish International Development Authority

Doff, A. 1990. *Teach English*. Cambridge: CUP

Dufon, M. 1993. Ethics in TESOL research. *TESOL Quarterly* 27.1: 157-60

Early, P. and Bolitho, R. 1981. Reasons to be cheerful: or helping teachers to get problems into perspective. In *Focus on the Teacher*. ELT Documents 110. The British Council: 71-85

Elley, W. 1989. Tailoring the evaluation to fit the context. In Johnson, R. (ed.) 1989: 270-285

Elliott, J. 1980. The evaluation of the dissemination stage of a Schools Council project. In McCabe, C. (ed.) 1980: 77-87

Estaire, S. 1993. The action plan cycle: a way of integrating the classroom and training course sessions. *The Teacher Trainer* 7.1: 26-28

Evans, M. and Hopkins, D. 1988. School climate and the psychological state of the individual as factors affecting the utilisation of educational ideas following an in-service course. *British Educational Research Journal* 14.3: 211-30

Festinger, L. and Katz, D. (eds) 1954. *Research Methods in the Behavioural Sciences*. London: Staples Press

Fox, G. 1980. *Reflecting upon Evaluation. Synthesis Report. In-Service Education and Training of Teachers: Towards New Policies*. Paris: Centre for Educational Research and Evaluation, OECD

Gaies, S. 1992. An approach to the evaluation of ELT preparation programs. In Sadtono, E. (ed.) 1992. *Language Teacher Education in a Fast-Changing World*. Anthology Series 29. Singapore: Regional Language Centre: 14-33

Gilroy, D. and Day, C. 1993. The erosion of INSET in England and Wales: analysis and proposals for a redefinition. *Journal of Education for Teaching* 19.2: 141-157

Green, C. 1993. When teachers write journals: a reflective approach. *Newsletter of IATEFL Special Interest Group for Teacher Training*. Nos. 8/9, Spring/Summer 1993: 4-6

Greenwood, J. 1985. Bangalore revisited: a reluctant complaint. *ELT Journal* 39.4: 268-73

Guba, E. and Lincoln, Y. 1981. *Effective Evaluation*. San Fancisco, Calif.: Jossey Bass

Guba, E. and Lincoln, Y. 1989. *Fourth Generation Evaluation*. Newbury Park: Calif.: Sage

Haill, A. 1990. Writing as a learning process in teacher education and development. *The Teacher Trainer* 4.1: 6-10

Hamilton, D., Jenkins, D., King, C., MacDonald, B. and Parlett, M. (eds) 1977. *Beyond the Numbers Game: a Reader in Educational Evaluation*. London: Macmillan

Hamp Lyons, E. 1986. Reaching the individual through open teachers' courses. Paper presented at IATEFL Conference, Brighton, April 1986.

Hargreaves, P. 1989. DES-IMPL-EVALU-IGN: an evaluator's checklist. In Johnson, R. (ed.) 1989: 35-47

Haworth, T. and Parker, R. 1995. The contribution of a face-to-face component in initial teacher training at a distance. In Howard, R. and McGrath, I. (eds): 78-94

Henderson, E. 1976. Attitude change in in-service training. *British Journal of In-Service Education* 2.2: 113-116

Henderson, E. 1978. *The Evaluation of In-Service Teacher Training*. London: Croom Helm

Henderson, E. 1980. The evaluation of short courses using a questionnaire. In McCabe, C. (ed.) 1980a: 62-68

Henderson, E. 1981. Teachers learning at a distance: alternative approaches and their alternative findings. In Salmon, A. (ed.) 1981: 76-101

Henderson, E., Perry, G. and Spencer, M. 1975. *The Coordination of In-service Training for Teachers*. Oxford: Department of Educational Studies, University of Oxford

Herman, J., Morris, L. and Fitz-Gibbon, C. 1987. *Evaluator's Handbook*. Beverly Hills, Calif.: Sage

Higgens, J. 1981. Course report: Teachers of English. University of Edinburgh: Institute for Applied Language Studies. Mimeo.

Hodgson, F. and Whalley, G. 1985. Evaluation of in-service education: the question of criteria. *British Journal of In-Service Education* 12.1: 44-47

Holliday, A. 1992. Tissue rejection and informal orders in ELT projects: collecting the right information. *Applied Linguistics* 13.4: 403-24

Holliday, A. 1994. Overseas projects: how do we find out what we need to know ? *IATEFL Special Interest Group Newsletter on Management* 16: 24-25

Hopkins, D. 1989. *Evaluation for School Development*. Milton Keynes: Open University

House, E. 1972. The conscience of educational evaluation. *Teachers College Record* 73: 405-414

Howard, R. and McGrath, I. (eds) 1995. *Distance Education for Language Teachers: a UK Perspective*. Clevedon, Avon: Multilingual Matters

Hundleby, S. and Breet, F. 1988. Using methodology notebooks on in-service teacher-training courses. *ELT Journal* 42.1: 34-36

Hyman, H. et al. 1954. *Interviewing in Social Research*. Chicago: University of Chicago Press

Institute for Applied Language Studies, University of Edinburgh. 1991. Assessment Guide. RSA Diploma in the Teaching of Foreign Languages to Adults (by distance learning) 1991-92. Edinburgh: IALS, University of Edinburgh. Mimeo.

Isaac, S. and Michael, W. 1981. *Handbook in Research and Evaluation*. San Diego, Calif.: EdITS

Jarvis, J. 1992. Using diaries for teacher reflection on in-service courses. *ELT Journal* 46.2: 133-143

Johnson, R. (ed.) 1989. *The Second Language Curriculum*. Cambridge: CUP

Joyce, B. and Showers, B. 1980. Improving in-service training: the messages of research. *Educational Leadership* 37: 379-85

Kadepurkar, H. 1997. Towards evaluation of trainer training: a case study of the English language development project in Maharashtra, India. In McGrath, I. (ed.) 1997: 196-204

Kahn, R. and Cannell, C. 1957. *The Dynamics of Interviewing*. New York: John Wiley

Keating, R. 1963. *A Study of the Effectiveness of Language Laboratories*. NY: Institute of Adminstrative Research, Teachers College

Kennedy, C. 1985. Formative evaluation as an indicator of students' wants and attitudes. *The ESP Journal* 4.2: 93-100

Kiely, R. and Murkowska, A. 1994. Initiating evaluation activities: insider and outsider perspectives. In Kiely, R., Murphy, D. and Rea-Dickins, P. (eds) 1994: 60-72

Kiely, R., Murphy, D. and Rea-Dickins, P. (eds) 1993. *Managing Change and Development ELT in ECE*. PRODESS Colloqium, 7-9 November 1993, Manchester. Manchester: The British Council

Krathwohl, D., Bloom, B. and Masia, B. 1964. *Taxonomy of Educational Objectives*. Handbook II: the Affective Domain. N.Y.: David McKay

Lamb, M. 1995. The consequences of INSET. *ELT Journal* 49.1: 72-80

Langley, P. 1987. *Doing Social Research*. Ormskirk, Lancs: Causeway Press

Lawrence, L. 1995. Using evaluation to improve teacher education programmes. In Rea-Dickins, P. and Lwaitama, A. (eds) 1995: 78-87

Lindeman, E. 1924. *Social Discovery*. New York: Republic

Long, M. 1984. Process and product in ESL program evaluation. *TESOL Quarterly* 18.3: 409-25

Lowe, T. 1987. An experiment in role reversal: teachers as language learners. *ELT Journal* 41.2: 89-96

MacDonald, B. 1976. Evaluation and the control of education. In Tawney, D. (ed.) 1976. *Curriculum Evaluation Today: Trends and Implications*. Schools Council Research Studies. Basingstoke: Macmillan: 125-136

Mackay, R. 1981. Accountability in ESP programs. *The ESP Journal* 1.2: 107-21

Mackay, R. 1988. Position paper: programme evaluation and quality control. *TESL Canada Journal* 5.2: 33-41

Mackay, R. 1994a. Undertaking ESL/EFL programme review for accountability and improvement. *ELT Journal* 48.2: 142-149

Mackay, R. 1994b. Making evaluation work. *PRODESS News 6*. Newsletter of the Project Development and Support Scheme. October 1994. London: The British Council/Thames Valley University: 5-6

Mackay, R. and Palmer, J. 1981. *Program Design and Evaluation*. Rowley, Mass.: Newbury House

Madaus, G., Stufflebeam, D. and Scriven, M. 1983a. Program evaluation: a historical overview. In Madaus *et al.* (eds) 1983: 3-22

Madaus, G., Stufflebeam, D. and Scriven, M. (eds) 1983b. *Evaluation Models*. Boston, MA: Kluwer-Nijhoff

Maingay, P. 1988. Observation for training, development or assessment. In Duff, T. (ed.) 1988. *Explorations in Teacher Training: Problems and Issues*. London: Longman: 118-131

Marsh, I. 1987. Teacher perceptions of an in-service course: you can't please all the people all of the time. *British Journal of In-Service Education* 14.1: 56-60

McCabe, C. (ed.) 1980a. *Evaluating In-Service Training for Teachers*. Windsor, UK: National Foundation for Educational Research

McCabe, C. 1980b. Aims and methods. In McCabe, C. (ed.) 1980: 14-27

McCabe, C. 1980c. What is an evaluator ? *British Journal of Teacher Education* 6.3: 196-202

McCabe, C. 1987.The external evaluator - inspector or management consultant *?* *Evaluation and Research in Education* 1.1: 1-7

McDonough, J. 1994. A teacher looks at teachers' diaries. *ELT Journal* 48.1: 57-65

McGinley, K. 1986. Coming to terms with evaluation. *System* 14.3: 335-341

McGrath, I. (ed.) 1997. *Learning to Train*. London: Prentice Hall

McGrath, I. 1982. The methodology component of a refresher course for teachers of English as a Foreign Language. M.Sc. dissertation proposal. University of Edinburgh: Department of Applied Linguistics

McGrath, I. 1995. Some thoughts on project evaluation. In British Council 1995: 15-20

McGrath, I. 1996. Participant action plans and the evaluation of teachers' courses. *Edinburgh Working Papers in Applied Linguistics*, Vol.7. Edinburgh: Department of Applied Linguistics/Institute for Applied Language Studies, University of Edinburgh: 85-99

McGrath, I. and Altay, S. 1990. Trainer training, simulation and the feedback imperative. *Edinburgh Working Papers in Applied Linguistics,* Vol. 1. Edinburgh: Department of Applied Linguistics/Institute for Applied Language Studies, University of Edinburgh: 35-47

McLaughlin, M. and Phillips, D. (eds) 1991. *Evaluation and Education at Quarter Century*. 9th Yearbook of the National Society for the Study of Education, Part 2. Chicago: University of Chicago Press

McLeish, J. 1969. *Teachers' Attitudes: a Study of National Differences*. Cambridge: Cambridge Institute of Education

Mitchell, R. 1991. Evaluation of foreign language teaching projects and programmes. In de Bok, K., Ginsberg, R. and Kramsch, C. (eds) 1991. *Foreign Language Research in Cross-Cultural Perspective*. Amsterdam: John Benjamins: 145-160

Morris, L. and Fitz-Gibbon, C. 1978. *How to Present an Evaluation Report*. Beverly Hills, Calif.: Sage

Morrow, K. and Schocker, M. 1993. Process evaluation in an INSET course. *ELT Journal* 47.1: 47-54

Mortimore, P. and Mortimore, J. 1989. School-focused in-service training in England and Wales: the challenge to higher education. *Journal of Education for Teaching* 15.2: 133-139

Munn, P. and Drever, E. 1990. *Using Questionnaires in Small-Scale Research. A Teachers' Guide*. Edinburgh: Scottish Council for Research in Education

Murphy, D. 1985. Evaluation in language teaching: assessment, accountability and awareness. In Alderson, J. (ed.) 1985b: 1-17

Murphy, D. 1994. Response to Theme 3: Evaluation. In Kiely, R., Murphy, D. and Rea-Dickins, P. (eds) 1994: 84-87

Murphy, D. 1995. Developing theory and practice in evaluation. In Rea-Dickins, P. and Lwaitama, A. (eds) 1995: 10-28

Murphy, D. 1996. The evaluator's apprentices. *Evaluation* 2.3: 321-338

Murphy-O'Dwyer, L. 1985. Diary studies as a method for evaluating teacher training. In Alderson, J. (ed.) 1985b: 97-128

Nevo, D. 1986. The conceptualisation of educational evaluation: an analytical review of the literature. *Review of Educational Research* 53.1: 117-28

Newstrom, J. and Scammell, E. 1980. *Games Trainers Play*. N.Y.: McGraw-Hill

Nisbet, J. and Watt, T. 1984. Case study. In Bell, J. *et al.* (eds) 1984: 72-92

Nunan, D. 1992. *Research Methods in Language Learning*. Cambridge: CUP

Oppenheim, A. 1966. *Questionnaire Design and Attitude Measurement*. London: Heinemann

Overseas Development Adminstration (ODA). 1984. *The Evaluation of Aid Projects and Programmes*. London: HMSO

Palmer, C. 1992. Diaries for self-assessment and INSET programme evaluation. *European Journal of Teacher Education* 15.3: 227-238

Palmer, G. 1992. The practical feasibility of diary studies for INSET. *European Journal of Teacher Education* 15.3: 239-254

Parker, R. and Graham, T. 1995. Monitoring the effectiveness of distance learning as a means of partially delivering teacher education. In Howard, R. and McGrath, I. (eds) 1995: 161-173

Parlett, M. and Hamilton, D. 1972. *Evaluation as Illumination: a New Approach to the Study of Innovatory Programmes*. Occasional Papers 9. Centre for Research in the Educational Sciences, University of Edinburgh

Parsons, C. 1990. The politics, ethics and diplomacy of reporting educational evaluation. *Evaluation and Research in Education* 4.3: 147-154

Patton, M. 1986 (2nd ed.) *Utilization-focused Evaluation*. Beverly Hills, Calif.: Sage

Patton, M. 1990. *Qualitative Evaluation and Research Methods*. Beverly Hills, Calif.: Sage

Peck, A. 1988. *Language Teachers at Work*. London: Prentice Hall

Peil, M. 1982. *Social Science Research Methods: an African Handbook*. London: Hodder and Stoughton

Perkins, K. and Angelis, P. 1985. Some considerations for ESL program evaluation. *RELC Journal* 16.2: 72-92

Porter, P., Goldstein, L., Leatherman, J. and Conrad, S. 1990. An ongoing dialogue: learning logs for teacher preparation. In Richards, J. and Nunan, D. (eds) 1990: 227-240

Potter, M. 1994. The way forward. *PRODESS News* 6. Newsletter of the Project Development and Support Scheme. October 1994. London: The British Council/Thames Valley University: 7

Potts, P. 1985. The role of evaluation in a communicative curriculum, and some consequences for materials design. In Alderson, J. (ed.) 1985a.: 19-79

Powney, J. and Watts, M. 1987. *Interviewing in Educational Research.* London: Routledge and Kegan Paul

Provus. M. 1971. *Educational Evaluation and Decision-Making*. Itasca, Ill.: Peacock

Rea, P. 1983. Evaluation of educational projects, with special reference to English language education. In Brumfit, C. (ed.) 1983. *Language Teaching Projects for the Third World.* ELT Documents 116. Oxford: Pergamon/British Council: 85-98

Rea-Dickins, P. 1991. The principles underlying the evaluation of training. In British Council. 1991. *Training for Sustainability of ELT Projects*. Dunford House Seminar Report 1990. London: British Council: 41-49

Rea-Dickins, P. 1994. Evaluation and English language teaching. State of the art article. *Language Teaching* 27.2: 71-91

Rea-Dickins, P. and Germaine, K. 1992. *Evaluation*. Oxford: OUP

Rea-Dickins, P. and Lwaitama, A. (eds) 1995. *Evaluation for Development in English Language Teaching*. Review of English Language Teaching 3.3 (1993). London: Macmillan

Reid, M. 1994. Training learners to be evaluators. *PRODESS News* 5. Newsletter of the Project Support Development Group. London: British Council/Thames Valley University: 6-7

Reynolds, M. 1982. The choice of frame and focus in a data-driven case study: a description of a research process. In Dingwall, S. and Mann, S. (eds) 1982: 40-58

Richards, J. and Nunan, D. (eds) 1992. *Second Language Teacher Education*. Cambridge: CUP

Richards, K. 1995. Direction and debate in distance materials for teacher development. In Howard, R. and McGrath, I. (eds) 1995: 142-160

Rinvolucri, M. 1983. Writing to your students. *ELT Journal* 37.1: 16-21

Rockwell, S. 1982. Evaluation abstract: a responsive approach to evaluating an alcohol program for youth. *Studies in Educational Evaluation* 8: 175-83

Ross, A., McNamara, D. and Whittaker, J. 1977. *An Experiment in Teacher Education*. Research into Higher Education Monograph 29. Guildford, Surrey: Society for Research into Higher Education, University of Surrey

Rowntree, D. 1985. *Developing Courses for Students*. London: Paul Chapman

Rudduck, J. 1981. *Making the Most of the Short In-service Course.* Schools Council Working Paper No.71. London: Methuen

Rudduck, R. 1981. *Evaluation: a Consideration of Principles and Methods*. Manchester Monographs 18. University of Manchester

Salmon, A. (ed.) 1981. *The Evaluation of In-Service Education and Training of Teachers.* Lisse: Swets and Zeitlinger (for the Council of Europe)

Sanders, J. 1992. *Evaluating School Programs: an Educator's Guide*. Newbury Park, Calif.: Corwin Press

Scherer, G. and Wertheimer, M. 1964. *A Psycholinguistic Experiment in Foreign Language Teaching*. New York: McGraw Hill

Scriven M. 1967. The methodology of evaluation. In Tyler, R., Gagne, R. and Scriven, M. 1967. *Perspectives of Curriculum Evaluation*. AERA Monograph Series on Curriculum Evaluation, No.1. Chicago: Rand McNally: 39-83

Scriven, M. 1974. Pros and cons about goal-free evaluation. *Evaluation Comment* 3: 1-4

Scriven, M. 1994. Evaluation as a discipline. *Studies in Educational Evaluation* 20: 147-166

Scriven, M. 1996. The theory behind practical evaluation. *Evaluation* 2.4: 393-404

Sharp, A. 1990. Staff/student participation in course evaluation: a procedure for improving course design. *ELT Journal* 44.2: 132-7

Sheal, P. 1989. *How to Develop and Present Staff Training Courses*. London: Kogan Page

Siedow, M., Memory, D., and Bristow, P. 1985. *Inservice Education for Content Area Teachers*. Newark, Del.: International Reading Association

Simons, H. 1979. Ethical principles in school self-evaluation. In Bell, J. *et al.* (eds) 1984: 123-132

Slembrouck, S. 1987. Questions and questionnaires; answers and answerraires. In Littlejohn, A. and Melouk, M. (eds) 1987. *Research Methods and Processes*. Lancaster: Department of Linguistics and Modern English Language,University of Lancaster: 76-90

Smith, P. Jr 1970. *A Comparison of the Cognitive and Audiolingual Approaches to Foreign Language Instruction: the Pennsylvania Foreign Language Project*. Philadelphia, Pennsylvania: Center for Curriculum Development

Sparrow, F. Foreword. In Rudduck, J. 1981: 9-10

Spelling, K. 1981. Introductory paper. In Salmon, A. (ed.) 1981: 3-14

Stake, R. 1976. *Evaluating Educational Programmes: the Need and the Response*. Paris: OECD

Stake, R. 1986. Evaluating educational programmes. In Hopkins, D. (ed.) 1986. *Inservice* Training and Educational Development. London: Croom Helm: 245-8

Stake, R. et al. 1975. *Evaluating the Arts in Education: a Responsive Approach*. Columbus, Ohio: Merrill

Stenhouse, L. 1975. *An Introduction to Curriculum Research and Development*. London: Heinemann

Stufflebeam, D., Foley, W., Gephart, W., Guba, E., Hammond, R., Merriman, H. and Provus, M. 1971. *Educational Evaluation and Decision-Making*. Itasca, Ill.: Peacock

Stufflebeam, D. 1990. Professional standards for educational evaluation. In Walberg, H. and Haertel, G. (eds) 1990. *Evaluating Education Performance: a Sourcebook of Methods, Instruments and Examples*. Berkeley, Calif: McCutchan: 94-106

Taylor, P. 1980. A review of some in-service evaluation studies undertaken in the UK. In McCabe, C. (ed.) 1980a: 28-38

TESOL Research Committee. 1983. Guidelines for Ethical Research in ESL. *TESOL Quarterly* 14.3: 383-88.

Thornbury, S. 1991. Watching the whites of their eyes: the use of teaching- practice logs. *English Language Teaching Journal* 45.2: 140-146

Thorpe, M. 1988. *Evaluating Open and Distance Learning*. London: Longman

Tisher, R. and Wideen, M. 1990. The role played by research in teacher education. In Wideen and Tisher (eds) 1990a: 1-10

Tribble, C. 1993. Professional development for teachers in the Baltic States. *PRODESS News* 3. London: British Council/Thames Valley University: 2-3

Tyler, R. 1949. Evaluating learning experiences. In Hamilton, D. et al.(eds) 1977: 29-34
Vance, C. 1979. Personal journals in training. *Training and Development Journal,* August 1979: 54-55
Vasconcelos, G. 1994. Trainee feedback. *The Teacher Trainer* 8.2: 13-14
Vorbeck, M. 1981. Preface to Salmon, A. (ed.) 1981.
Walker, R. 1985. *Doing Research.* London: Methuen
Wallace, M. 1991. *Training Foreign Language Teachers. A Reflective Approach.* Cambridge: Cambridge University Press
Wallace, M. 1997. Designing courses for trainers. In McGrath, I. (ed.) 1997: 174-185
Ward, T., Barr, B., Chai, M., Hua, D., Kong, X. and Lu, H. 1995. *A Crane Among the Chickens? Evaluating a Training Programme for English Teachers in China.* Shanghai: Shanghai Foreign Language Education Press
Weir, C. and Roberts, J. 1994. *Evaluation in ELT.* Oxford: Blackwell
Weis, D. *et al.* 1974. Dogmatism and the inner-city teacher. *Journal of Studies of the Personnel Association* 13.1: 2-6
Wideen, M. and Tisher, R. (eds) 1990a. *Research in Teacher Education: International Perspectives.* Basingstoke, Hants: Falmer
Wideen, M. and Tisher, R. 1990b. Review, reflections and recommendations. In Wideen, M. and Tisher, R. (eds) 1990a: 255-267
Wise, R. 1980. The evaluator as educator. *New Directions for Program Evaluation* 5: 11-18
Woodward, T. 1988. *Loop Input.* Canterbury: Pilgrims
Woodward, T. 1991. *Models and Metaphors in Language Teacher Training.* Cambridge: CUP
Woodward, T. 1992. *Ways of Training.* London: Longman
Wragg, E. 1984. Conducting and analysing interviews. In Bell, J. *et al* (eds) 1984: 177-197
Youngman, M. 1984. Designing questionnaires. In Bell, J. *et al* (eds) 1984: 156-176

APPENDIX 1.1

Evaluation models: a selective account

The professionalisation of educational evaluation

The beginning of a decade is often marked by retrospection, and as far as educational evaluation is concerned, the early 1990s were no exception. 'Evaluation and Education: at Quarter Century' (McLaughlin and Phillips 1991), for instance, contains updated position papers by such key figures as Tyler, Eisner, Scriven, Stake and Stufflebeam.

What is clear from these accounts is that although concern about evaluation found expression from 1930s onwards, notably in the work of Tyler, it was not until the 1960s that educational evaluation emerged as a professional field. It now has a code of practice, pressed for by Cronbach in the 1970s, but only deemed to be appropriate in 1981 (Standards for Evaluation of Educational Programs, Projects and Materials), training programmes for evaluators (in North America at least), and fora for professional exchange in the form of associations (e.g. American Evaluation Association 1985; UK Evaluation Association 1994) and journals dedicated to evaluation (e.g. *Studies in Educational Evaluation*, *Evaluation Review*, *Evaluation Newsletter* and most recently *Evaluation*).

A brief historical overview

Madaus, Stufflebeam and Scriven's (1983a) historical overview of the development of educational evaluation distinguishes six ages: (1) pre-1900 (the Age of Reform) (2) 1900-1930 (the Age of Efficiency and Testing) (3) 1930-1945 (the Tylerian Age) (4) 1945-1957 (the Age of Innocence - or Ignorance) (5) 1958-72 (the Age of Expansion) and (6) 1973- (the Age of Professionalization), the final age being characterised by the development of professional standards, associations and forms of professional exchange (journals, etc).

Programme evaluation, as distinct from performance evaluation (learner assessment), dates only from the third of these ages, that named after Ralph Tyler, who coined the term 'educational evaluation' (to refer to the assessment of objectives-achievement) and is often referred to as the father of educational evaluation (Madaus *et al, op.cit.*: 8). Tyler established a model of programme evaluation where none had existed; this model was extremely influential for a time; and when its inadequacies were revealed in due course this forced others to develop alternative models.

One way of distinguishing between these models is by what Guba and Lincoln (1981) refer to as their 'organiser'. Tyler's approach is organised according to *objectives*. The content of a programme is described (by means of a carefully specified set of procedures) in terms of a set of behavioural objectives and student success in achieving these objectives is equated with programme success. It was 'a rational alternative to the norm-referenced testing that preceded it' (Guba and Lincoln 1982: 6) and it appealed to teachers, 'who testified eagerly to its utility in revealing their previously hidden assumptions, in providing feedback, and in forcing them to think explicitly about what they were trying to do' (*ibid*). Tyler had his followers (e.g Popham in the USA and, as Hopkins (1989) notes, Wiseman

and Pidgeon (1972) in the UK), but the objectives-based model had a number of weaknesses. One problem is with objectives as an organiser. For instance, the approach does not help us to understand the educational process; the objectives are not the result of negotiation between those involved in the process; and the necessity to render educational purposes in behavioural terms can result in their trivialisation or neglect (Stenhouse 1975). Moreover, as Guba and Lincoln (*op.cit.*) point out, the objectives were not themselves subject to evaluation; and nor were standards provided on which to base decisions relating to curriculum change.

In the late 1950s America experienced what Hopkins (1989: 6) refers to as a 'post-sputnik flurry of curriculum development', and a massive increase in large-scale evaluation associated with this development. Guba and Lincoln make the connections more forcefully:

> On October 4, 1957, the Russians launched Sputnik and changed the face of American education. Inadequate schooling was blamed for the fact that America was now Number 2 - and at once the public determined to try harder. Translated into action, trying harder meant refurbishing the curricula of the schools. Millions of dollars were suddenly poured into the development of new courses ... Of course, if such massive resources were to be expended, it was essential that the resulting products be evaluated.
>
> (Guba and Lincoln 1982: 7)

Recognition of the need for social reform in the 1960s also resulted in new programmes, with associated evaluation.

Underpinning this massive expansion were a number of developments. The Educational Testing Service had been founded in 1947 by E.F. Lindquist, Tyler and others and this had led to the growth in standardised testing. There had also been advances on other fronts: technology for scoring and analysing tests; taxonomies of objectives (Bloom, Engelhart, Furst, Hill and Krathwohl 1956; Krathwohl, Bloom and Masia 1964); and experimental design (Madaus *et al, op.cit.*).

Herman, Morris and Fitz-Gibbon's (1987) cogent historical summary picks up the story at this point, i.e. periods 5 and 6 in Madaus *et al*'s chronology. Within these periods they observe three trends.

> Models advanced in the late 1960s and early 1970s were fueled largely by the needs of large-scale curriculum developers who needed formative information to direct their revision efforts, by those of their sponsors who wanted comparative summative information to guide funding decisions, and by federal policymakers who wanted to ensure acountability for their social reforms. These models emphasized experimental methods, standardized data collection, large samples, and the provision of scientific, technical data; they also reflected general optimism that systematic, scientific measurement procedures would deliver unequivocal evidence of program success or failure. "Hard data" (i.e. empirically based data) would, it was hoped, provide both sound information for planning more effective programs and a rational basis for educational, social service, and other policy decision making. Clear cause-effect relationships, it was assumed, could be established between programs and their outcomes, and program variables could be manipulated to reach desired effects. This first wave of models - based initially on

goals and then later on decisions, issues or problems - provided decision makers and policy-makers with aggregated, standardized, and technically rigorous information about prespecified program processes and outcomes. Some accused these models, however, of being superficial and insensitive to important variations in local programs.

Following this first, quantitatively grounded wave, a second wave emerged. These models were characterized by a responsiveness to unique characteristics and processes within local settings and to issues as perceived by stakeholders. Rather than assuming that social programs were discrete and easily prespecified in terms of process and outcomes, these models acknowledged that social and other programs often are complex, amorphous mobilizations of human activities and resources that vary significantly from one locale to another, embedded in and influenced by complex political and social networks. Rare is the program, according to these model builders, which exists in hermetically sealed isolation, perfectly appropriate for scientific measurement and duplication. Their models, as a result, stressed the importance of naturalistic, qualitative methods for understanding the means of operation and the effects of programs. While providing in-depth understanding of unique program configurations - understanding that is critical to program improvement aims - these models unfortunately lacked easy or credible ways of aggregating or generalizing findings across sites, a distinct disadvantage for accountability and/or higher-level decision-making.

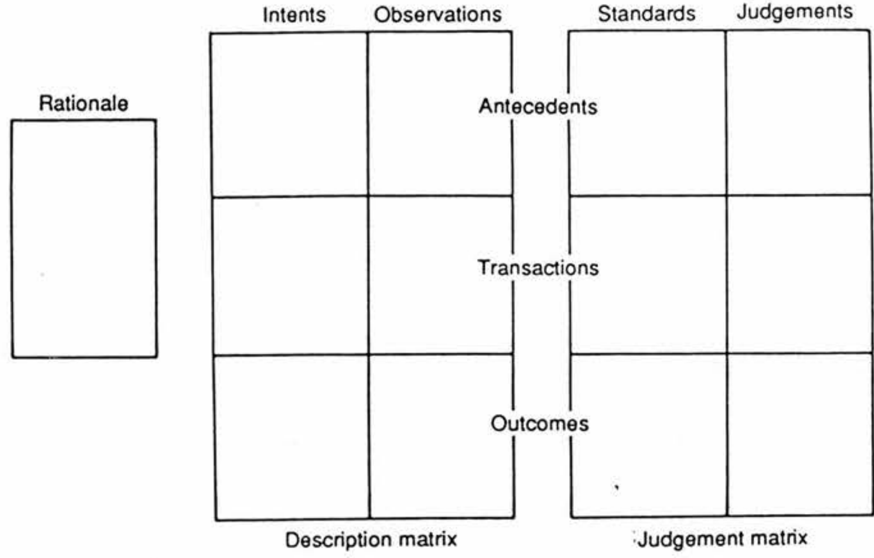
Concurrent with the growth of responsive models came renewed attention to the utility of evaluation findings. Spurred by federal and state mandates in the late 1960s and 1970s, thousands of evaluations were conducted across the country. Unfortunately many, if not most, of these evaluations did not have the expected impact; researchers questioned whether these evaluations had any impact at all. Continuing to believe in the potential contribution of their work to social policy, planning, and practice, some evaluators became concerned about how to ensure that their findings were used, not simply filed. These model builders stressed the importance of socio-political and other factors beyond technical quality which were critical to an effective (i.e. useful) evaluation process.

(Herman *et al.*: 1987: 9)

Herman *et al*'s tabulation of the 'the most prominent' models, their emphases, and associated readings appears as Appendix 1.2. There follow descriptions of a number of the best known of these.

A selection of models

Robert Stake's **'countenance model'**, which takes its name from the title of his paper 'The countenance of evaluation' (1967) and its separation of the two 'countenances' of description and judgement (see below), is a refined version of Tyler's approach. Although it also uses objectives as an organiser, it takes a broader view of these, relating 'transactions' during instruction to prior and future behaviour; it incorporates an observational element; and it evokes the notion of standards, relative and absolute.



Stake (1967), reproduced in Hopkins (1989: 26)

The drawbacks of the model were that it was not easy to use; the means for deriving standards were largely unspecified; no guidance was given on where to find or how to take account of unintended effects (though their possible existence was at least acknowledged); and there was no recognition of the possibility of conflicting values (Guba and Lincoln, *op.cit.*).

Critics of the objectives-based approach in America during the 1960s include Lee Cronbach (1963) and Malcolm Scriven (1967).

Cronbach (*op.cit.*) was concerned at the reliance on standardised testing and outcomes assessment. Defining evaluation as 'the collection and use of information to make decisions about an educational program' (1963/83: 101-2), he argued that for purposes of course revision information was needed on decision-making processes during a programme. Although this argument appears to have had little effect at the time, it was to have an influence on later models.

Scriven's paper, 'The methodology of evaluation', which draws a number of key distinctions, was another seminal contribution. Scriven distinguishes between formative and summative evaluation (i.e. between improving and judging the 'evaluand'); between goal-achievement and evaluation (the latter including assessment of the goals themselves); and between process and outcome evaluation. Scriven went on to develop the **'goal-free model'** (Scriven 1974), which takes effects as its organiser (Guba and Lincoln, *op.cit.*)). The goal-free model represents an almost perverse from of opposition to the objectives-based model. The premise is that if an evaluator knows what the objectives of a programme are, s/he is likely to ignore or not look particularly hard for other effects. Scriven therefore proposes that the evaluator should try to remain ignorant of intended effects and instead look for actual effects. Once found, these can be related to needs and the success of the programme assessed on the basis of the extent to which needs are met. Among the problems with the proposal are the fact that it gives no guidance on what kinds of effect to look for or how standards for judgement can be arrived at (Guba and Lincoln, *op.cit.*). Scriven later acknowledged that it would be preferable to operate goal-free and goal-based evaluation in tandem (Guba and Lincoln, *op.cit.*).

Cronbach's influence can be seen in the **CIPP (Context - Input - Process - Product) model**, proposed by Daniel Stufflebeam and associates (1971). The intention of the model, as Guba and Lincoln (*op.cit.*) note, was to generate a taxonomy of decision types, each of which could be serviced by an evaluation type (see below):

	Context Evaluation	*Input Evaluation*	*Process Evaluation*	*Product Evaluation*
Objective	To define the institutional context, to identify the target population & assess their needs, to identify opportunities for addressing the needs, to diagnose *problems* underlying the *needs*, & to judge whether proposed objectives are sufficiently responsive to the assessed needs.	To identify & assess *system capabilities*, alternative program *strategies*, procedural *designs* for implementing the strategies, budgets, & schedules.	To identify or predict, in process, *defects* in the procedural design or its implementation, to provide information for the preprogramed decisions, and to record & judge procedural events & activities.	To collect descriptions & judgments of outcomes & to relate them to objectives & to context, input, & process information; & to interpret their worth & merit.
Method	By using such methods as system analysis, survey, document review, hearings, interviews, diagnostic tests, & the Delphi technique.	By inventorying & analyzing available human & material resources, solution strategies, & procedural designs for relevance, feasibility & economy; and by using such methods as literature search, visits to exemplary programs, advocate teams, & pilot trials.	By monitoring the activity's potential procedural barriers & remaining alert to unanticipated ones, by obtaining specified information for programed decisions, by describing the actual process, & by continually interacting with & observing the activities of project staff.	By defining operationally & measuring outcome criteria, by collecting judgments of outcomes from stakeholders, & by performing both qualitative & quantitative analyses.
Relation to Decision-making in the Change Process	For deciding upon the *setting* to be served, the *goals* associated with meeting needs or using opportunities, & the *objectives* associated with solving problems, i.e., for *planning* needed changes; and for providing a basis for judging outcomes.	For selecting *sources of support*, solution *strategies*, & procedural *designs*, i.e., for *structuring* change activities; and to provide a basis for judging implementation.	For *implementing and refining the program design and procedure*, i.e., for effecting *process control*; & to provide a log of the actual process for later use in interpreting outcomes.	For deciding to *continue*, *terminate*, *modify*, *or refocus* a change activity, & present a clear record of effects (intended & unintended, positive & negative).

Stufflebeam, D. 1983. Four types of evaluation. From 'The CIPP model for program evaluation'. In Madaus *et al.* (eds) 1983: 129

Apart from its clarity, the virtue of the model is that it recognises the importance of contextual and process information. However, it makes certain assumptions (e.g.about the transparency of the decision-making process or the ease with which decision-makers and decision-making processes can be identified) which are not necessarily justified (Guba and Lincoln, *op.cit.*).

The major UK contribution, **illuminative** evaluation, also focuses on context and process. Malcolm Parlett and David Hamilton's (1972) well known paper 'Evaluation as illumination: a new approach to the study of innovatory programmes' argues against what Parlett and Hamilton call the 'agricultural-botany' approach based on hypotheses, experimental design and statistical analysis and in favour of an approach which 'illuminates' by offering a description and interpretation of the context within which a programme takes place. Interest in this paradigm shift, from a scientific to a naturalistic approach drawing on the methods of anthropology, ethnography, and sociological field, led to a number of conferences in Cambridge and culminated in a collection published in 1977 and tellingly titled 'Beyond the Numbers Game' (Hamilton, Jenkins, King, MacDonald and Parlett).

In **'responsive' evaluation**, a term first used by Stake (1975), the organiser is the concerns and issues voiced by the 'stakeholding' audiences (also Stake's term). As well as having different concerns, these audiences are also likely to have different values and to have different information needs. It follows that feedback to them should differ accordingly. In attempting to identify the various concerns and issues and respond appropriately to these audiences, the evaluator also adopts a role markedly different from that which is customary: interacting with those affected rather than standing outside and looking in. The four phases in responsive evaluation are described by Guba and Lincoln (*op.cit.*) as follows: (1) initiating and organising evaluation - involving identification of stakeholders and negotiation of purpose, foci and rights of access to records (2) identifying (through interviews and questionnaires) concerns, issues and values of stakeholders (3) obtaining information relevant to (2) through e.g. observation, interviews, questionnaires and tests (4) preparing reports (often using a case study format). Key differences in the approach employed, compared with the scientific paradigm, are the absence of a prior research design and the use of naturalistic data collection methods.

In distinguishing between the various models (and the above is only a selection of the best-known) Guba and Lincoln's 'organiser' principle is helpful, but other attempts have of course been made to show the similarities and differences. Stake (1986), for instance, contains an overview of the most common dimensions of evaluation designs (Appendix 1.3) and Nevo (1986), attempting a conceptualisation of the state of the art of educational evaluation, suggested ten questions to represent 'the major issues addressed by the most prominent evaluation approaches in education' (Nevo 1986: 15, cited in Hopkins 1989: 14). Nevo's questions are reproduced as Appendix 1.4. In a more recent comparison of what he calls six 'views' of programme evaluation, Scriven (1994) draws particular attention to the role of the evaluator in making explicit judgements. The six views are summarised below:

A. the 'strong decision support' view: evaluative conclusions are intended to assist the decision-maker (typically for purposes of accountability). Goal-achievement is one concern (as in the Tylerian approach) but consideration may also be given to the match between goals and needs (CIPP model).

B. the 'weak decision support' view: the evaluator makes no judgements, simply presenting data which can be used by the decision-maker to draw evaluative conclusions (the position adopted by Alkin 1991)

C. the 'relativistic' view: the evaluator works within the field of reference and values of the client, observing for instance any discrepancy between the intended and actual programme ('discrepancy' evaluation is associated with Provus 1971)

D. the 'rich description' approach; the evaluator reports what is observed, again without any explicit judgement (associated with Stake and UK evaluators of the 'illuminative' school)

E. the 'social process' school: the evaluator seeks to understand what is going on in the programme in order to improve it; this is seen as of considerably more importance than summative evaluation (associated with Cronbach and associates)

F. the 'constructivist' or 'fourth generation' approach: the evaluator sees evaluation not as an attempt to establish merit or worth but truth, which is constructed by individuals and negotiation within groups (Guba and Lincoln 1989).

Scriven's own view, and it is, as he observes, the commonsense view, is that it is the responsibility of evaluators to make judgements; this should not, however, be confused with making recommendations, which are in many cases best left to the decision-maker with inside knowledge of possibilities and constraints. (This stricture need not therefore apply to evaluation carried out by insiders, though the relationship between evidence and recommendations still needs to be established.)

Seen simply as a way of raising awareness, the models are interesting and useful. As a guide to action, especially for the part-time evaluator, they have their limitations.

APPENDIX 1.2

Some models of programme evaluation

Model	Emphasis	Selected References
Goal-oriented evaluation	Evaluation should assess student progress and the effectiveness of educational innovations.	Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill. Popham, W. J. (1975). *Educational evaluation*. Englewood Cliffs, NJ: Prentice-Hall.
Decision-oriented evaluation	Evaluation should facilitate intelligent judgments by decision makers	Stufflebeam, D. L. (Ed.). (1971). *Educational evaluation and decision-making*. Itasca, IL: F. E. Peacock. Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment, 2*, 2-7.
Responsive evaluation	Evaluation should depict program processes and the value perspectives of key people.	Stake, R. E., et al. (1975). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Charles E. Merril.
Evaluation research	Evaluation should focus on explaining effects, identifying causes of effects, and generating generalizations about program effectiveness.	Campbell, D. (1969). Reforms as experiments. *American Psychologist, 24*, 409-429. Rossi, A., & Freeman, H. E. (1985). *Evaluation: A systematic approach* (3rd ed.). Newbury Park, CA: Sage.
Goal-free evaluation	Evaluation should assess program effects based on criteria apart from the program's own conceptual framework, especially on the extent to which real client needs are met.	Scriven, M. (1974). Pros and cons about goal-free evaluation. In W. J. Popham (Ed.), *Evaluation in education: Current applications*. Berkeley, CA: McCutchan.
Advocacy-adversary evaluation	Evaluation should derive from the argumentation of contrasting points of view.	Wolf, R. L. (1975). Trial by jury: A new evaluation method. *Phi Delta Kappan* (November).
Utilization-oriented	Evaluation should be structured to maximize the utilization of its findings by specific stakeholders	Patton, M. Q. (1986). *Utilization-focused evaluation*. Newbury Park, CA: Sage.

Herman *et al.* 1987: 10

APPENDIX 1.3

Dimensions of evaluation

Formative–summative. The most pervading distinction is the one between evaluation studies done during the development of a programme and those done after the programme has been completed. It is difficult to distinguish between the summative evaluation of a completed component and the formative evaluation of a part of the programme. The distinction is not clear-cut. The most useful distinction here may be between the users of the evaluation findings. Elsewhere, I have noted that when the cook tastes the soup it is formative evaluation and when the guest tastes the soup it is summative. The key is not so much when as why. What is the information for, for further preparation and correction or for savouring and consumption? Both lead to decision making, but toward different decisions.

Formal–Informal. Informal evaluation is a universal and abiding human act, scarcely separable from thinking and feeling. Formal evaluation is more operationalized and open to view, and less personal. It is needed when the results are to be communicated elsewhere. Of the two, the formal evaluation study is under an obligation to pass tests of accuracy, validity, credibility and utility.

Case particular–generalization. A most important distinction is between the study of a programme as a fixed and ultimate target, or the study of a programme as a representative of others. Most research is expected to be generalized in some ways: over time, over settings or over subject matters, for example. Evaluation research may be done essentially to discover the worth of the particular programme, or the worth of the general approach. Studies are perceived very differently in this regard, both by investigators and their audiences; and a large misperception is possible. The more the study is expected to be a basis for generalization, the more the need for controls, controlled variation, or careful description of uncontrolled variation.

Product–process. Another dimension on which evaluation studies vary is as to whether they give primary attention to the outcomes of the programme or to its transactions. A study of the 'product' is expected to indicate the pay-off value; a study of the 'process' is expected to indicate the intrinsic values of the programme. Both are needed in any effort to get at a full indication of the worth of the programme, but in any actual study only a small portion of either can be examined.

Descriptive–judgemental. many evaluators coming from a social science background define the evaluation task largely as one of providing information, with an emphasis on objective data and a de-emphasis on subjective data. Those coming from the humanities are likely to reverse the emphases. One will find some studies highly descriptive of students and settings, providing careful reports of differences and correlations, but with little direct reference to criteria of worth and value standards. And elsewhere one will find evaluation studies probing into the pluralism of values to be found in any educational setting. As with any of these dimensions, any particular study is not likely to be at one pole or the other, but to make some combination the compromise.

Preordinate–responsive. Studies differ considerably as to how much the issues of evaluation are determined by observation of activities and by the realization of concerns of participants in the programme. Preordinate studies are more oriented to objectives, hypotheses and prior expectations, mediated by the abstractions of language. Preordinate evaluators know what they are looking for and design the study so as to find it. Responsive studies are organized around phenomena encountered – often unexpectedly – as the programme goes along.

Wholistic–analytic. Studies differ also as to how much they treat the programme as a totality, recognizing conceptual boundaries common to non-technical audiences. The more common social science research approach is to concentrate on a small number of key characteristics. A case study is often used to preserve the complexity of the programme as a whole, whereas a multivariate analysis is more likely to indicate the relationship among descriptive variables.

Internal–external. An obviously important difference in evaluation studies is whether they will be conducted by personnel of the institution responsible for the programme or by outsiders. They differ as to how formal the agreement to evaluate, as to how free the evaluators are to raise issues and interpret findings, and as to how changes in plans will be negotiated.

Stake (1986: 245-8), reproduced in Hopkins (1989: 16-18)

APPENDIX 1.4

Towards a conceptualisation of evaluation

1 *How is evaluation defined?*

Educational evaluation is a systematic description of educational objects and/or an assessment of their merit or worth.

2 *What are the functions of evaluation?*

Educational evaluation can serve four different functions: (a) formative (for improvement); (b) summative (for selection and accountability); (c) sociopolitical (to motivate and gain public support); and (d) administrative (to exercise authority).

3 *What are the objects of evaluation?*

Any entity can be an evaluation object. Typical evaluation objects in education are students, educational and administrative personnel, curricula, instructional materials, programs, projects, and institutions.

4 *What kinds of information should be collected regarding each object?*

Four groups of variables should be considered regarding each object. They focus on (a) the goals of the object; (b) its strategies and plans; (c) its process of implementation; and (d) its outcomes and impacts.

5 *What criteria should be used to judge the merit of an object?*

The following criteria should be considered in judging the merit or worth of an educational object: (a) responding to identified needs of actual and potential clients; (b) achieving national goals, ideals, or social values; (c) meeting agreed-upon standards and norms; (d) outdoing alternative objects; and (e) achieving important stated goals of the objects. Multiple criteria should be used for any object.

6 *Who should be served by an evaluation?*

Evaluation should serve the information needs of all actual and potential parties interested in the evaluation object ('stakeholders'). It is the responsibility of the evaluator(s) to delineate the stakeholders of an evaluation and to identify or project their information needs.

7 *What is the process of doing an evaluation?*

Regardless of its method of enquiry, an evaluation process should include the following three activities: (a) focusing the evaluation problem; (b) collecting and analyzing empirical data; and (c) communicating findings to evaluation audiences. There is more than one appropriate sequence for implementing these activities, and any such sequence can (and sometimes should) be repeated several times during the life span of an evaluation study.

8 *What methods of enquiry should be used in evaluation?*

Being a complex task, evaluation needs to mobilize many alternative methods of enquiry from the behavioural sciences and related fields of study and utilize them according to the nature of a specific evaluation problem. At the present state of the art, an *a priori* preference for any specific method of enquiry is not warranted.

9 *Who should do evaluation?*

Evaluation should be conducted by individuals or teams possessing (a) extensive competencies in research methodology and other data analysis techniques; (b) understanding of the social context and the unique substance of the evaluation object; (c) the ability to maintain correct human relations and to develop rapport with individuals and groups involved in the evaluation; and (d) a conceptual framework to integrate the above-mentioned capabilities.

10 *By what standards should evaluation be judged?*

Evaluation should strike for an optimal balance in meeting standards of (a) utility (to be useful and practical); (b) accuracy (to be technically adequate); (c) feasibility (to be realistic and prudent); and (d) propriety (to be conducted legally and ethically). (Nevo, 1986: 24–6)

Nevo (1986: 15), reproduced in Hopkins (1989: 14-16)

APPENDIX 1.5

Evaluation studies of INSET programmes for language teachers

AUTHOR	DATE	LOCATION	PROGRAMME TYPE/ PARTICIPANTS	FOCI	POINT OF EVALUATION	TECHNIQUES
Alderson	1985	UK	10 weeks; overseas teachers of ESP; award bearing	participant perceptions of value of course	follow-up	questionnaire, implementation plan, incidental contact with ex-participants and colleagues
Cumaranatunge	1989	Sri Lanka	1. 9 days (60 hours); for prospective local trainers of untrained teachers; 36 particpants	analysis of needs and draft course designs		critical discussion
				participant attendance, administrative arrangements, participant reactions to course		informal monitoring
				achievement of programme objectives		questionnaire
			2. 6 days (40 hours); for prospective local trainers of primary teachers; 20 participants	analysis of needs and draft course designs		critical discussion
				participant attendance; administrative arrangements; participant reactions to course		informal monitoring

			3. 10 weeks; for prospective Training College lecturers; 19 participants	participant attitudes to course and recommendations analysis of needs and draft course designs attendance, administrative arrangements participant self-assessment and reactions to specific aspects of course, aspects of course not dealt with in questionnaire, participant responses to questionnaire and group discussion, course design		questionnaire critical discussion informal monitoring, staff meetings questionnaire group discussion staff discussion informal external evaluation
Lamb	1995	Indonesia	10 days (25 hours); 16 tertiary level teachers	programme impact (methodology)	follow-up	interview, observation
McGrath & Altay	1990	UK	component of 1-year Master's	participant attitudes	during programme	simulation
Morrow & Schocker	1993	UK	3-week British Council Summer School; 70 participants	participant perceptions of/reactions to course content and methodology	during programme: - daily - weekly - end of course	individual interviews, staff meetings; posters, meetings with participants; questionnaires

Murphy-O'Dwyer	1985	UK	2-week course for French secondary school teachers; 15 participants	participant attitudes to learning affective factors; factors affecting participation	during programme follow -up	participant diaries individual interviews written retrospective comments by participants
Palmer, C.	1992	UK	3 short courses (2-3 weeks) for Norwegian secondary school teachers	participant reactions to the course	start of course, during, end during	questionnaires diaries (participants, tutors, observers)
Palmer, G.	1992	UK	2 short courses (2-3 weeks) for Norwegian secondary school teachers'; 31 participants	participant reactions to course	during course end of course post course follow-up	daily tutors' meetings, diaries (participants and tutors) questionnaire, discussion with participants participant reports to Ministry questionnaire on diary keeping

Richards	1995	UK/Japan	component of 2-year distance learning Master's	participant reactions to component materials	during component	informal feedback, meeting with participants
					end of component	component feedback form, unstructured interview, detailed written feedback on extracts, tutor analysis of materials
Wallace	in press	UK	3-week British Council summer course for teacher trainers	tutor/participant feelings about each component	end of component	tutor meeting with 'participant tutors' present
				participant reactions	mid-course	participant group meetings followed by oral reports to tutors
				participant reactions to course	end-of-course	participant feedback, oral and written

Ward *et al*	1995	China	1-year programme for lecturers in HE (143 participants 1984-88)	programme impact (methodology; classroom use of L1) participants' views of gains (e.g. language vs methodology)	follow-up	interview, observation, questionnaire
			2-year programme for senior middle school teachers (170 participants 1988-1992)	participant attitudes to methodological issues participant reactions to course programme impact (methodology; classroom use of L1)	during course: - beginning/ end of course - end of semester follow-up	questionnaire evaluation sheets interviews and informal conversations with participants, colleagues and students; observation (recorded) of participants and colleagues; questionnaire
Weir & Roberts	1994	Nepal	baseline study of 22 teachers (11 each in experimental and control groups) from 900 involved in 4-week programme	teacher language improvement; methodology student gains	follow-up	teacher self-report, observation language tests (students), students' work

APPENDIX 2.1

Application of an evaluation checklist

Target audience	Purpose	Focus	Criteria	Method	Means/ instruments	Agents	Resources	Time factors	Findings	Presentation of results	Follow-up
Non-specialist e.g. clients, sponsors, policy makers, resource allocators, etc.	*Formative* e.g. to revise, modify, monitor, assess progress/ need for adjustments, etc.	*Direct/ indirect* e.g. programme relevance; social impact; change in e.g. teacher behaviour, perceptions, etc. student enrolment, proficiency, performance, exam results, etc. affective factors such as motivation, attitude, etc.	*Global/ 'absolute'* adequacy/ success judged without reference to local constraints, etc.	*A priori* e.g. scrutiny by 'experts', etc.	*A priori* e.g. analytical procedures, etc.	*Internal* e.g. by self (i.e. where targets and agents are same); by one set of people involved with a programme/ project for another set of people involved with it, e.g. programme designer for resource allocators	*Staffing/ funding* e.g. - agents, people needed for 'means' e.g. testers, data collectors and analysers, etc. (with responsibilities spelt out and time allocated); - adequate funds to cover costs	*Timing and timescales* e.g. timing of formative and summative evaluation; adequate time allowance for collection of data, etc. built into planning	*Nature/ status of findings* e.g. advisory, mandatory, etc.	*Format* e.g. report; recommend-ations; etc.	*Action* e.g. what (in general/in detail)? by whom? by when? etc.
Specialist e.g. designers of programmes, curricula, syllabuses, courses, materials; teachers; students; etc.	*Summative* e.g. to justify change/ investment, etc. (Accountability)		*Relative* pre-determined/ agreed criteria according to local/ immediate requirements, needs, etc.	*Empirical* comparison of situation before implementation with situation during and/or after implementation	*Empirical* collection of base-line data and subsequent comparison - quantitative study: measurement via tests, questionnaires, observation, etc. - qualitative study: description via case-study, diary-study, observation, interview, etc.	*External* 'independent' evaluators brought in from outside					

Hargreaves 1989: 46

APPENDIX 2.2

ODA project framework

(and illustration of this approach)

	Teacher training/ trainer training	**Indicators of achievement**	**How assessed**	**Assumptions**
WIDER OBJECTIVES	To improve the standard of English amongst university entrants to contribute to Nigeria's development in science and technology.	Improved levels at university entrance. Improved levels of subject attainment through better English during university course.	University entrance exam results in English. Exam results and survey of subject specialist departments.	English remains language of instruction. Universities function. exams and entrance requirements remain constant. Coorperation of departments on evaluation.
IMMEDIATE OBJECTIVES	To improve in-service training of English teachers at junior secondary school level.	Numbers of courses run and numbers of teachers attending. Competent training. Competent teaching. Improved levels of English at end of junior secondary school.	Statistics from inspectorate, etc. Observation and report, evaluation by trainees (*sic*). Observation by trainers, inspectors, etc. Exam results at junior secondary school.	Trainers stay in jobs. Trainers initiate. Resources provided and continued. Incentives provided. Trainees can attend. Evaluation valid.
OUPUTS	A cadre of trained teacher trainers.		Project reports including end-of-course assessment.	Trainees turn up. Resources provided. Existence of local administrative structures.
INPUTS	Consultancy, workshops, materials, personnel, training.		Project reports.	Various parties agree. Recruitment of ELTOs. Trainers stay in jobs. Resources provided and continued. Incentives provided. Trainees can attend.

based on Weir and Roberts 1994 : 37-39

APPENDIX 2.3

Standards for evaluations of educational programs, projects and materials

(Joint Committee 1981, summarised in Stufflebeam D. 1990)

(A) **Utility Standards**

The utility standards are intended to ensure that an evaluation will serve the practical information needs of given audiences. These standards are:

(A1) Audience Identification
Audiences involved in or affected by the evaluation should be identified, so that their needs can be addressed.

(A2) Evaluator Credibility
The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.

(A3) Information Scope and Selection
Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.

(A4) Valuation Interpretation
The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

(A5) Report Clarity
The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

(A6) Report Dissemination
Evaluation findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.

(A7) Report Timeliness
Release of reports should be timely, so that audiences can best use the reported information.

(A8) Evaluation Impact
Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

(B) **Feasibility Standards**

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal. These standards are:

(B1) Practical Procedures
The evaluation procedures should be practical, so that disruption is kept to a minimum, and that needed information can be obtained.

(B2) Political Viability
The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

(B3) Cost Effectiveness
The evaluation should produce information of sufficient value to justify the resources expended.

(C) Propriety Standards

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results. These standards are:

(C1) Formal Obligation

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

(C2) Conflict of Interest

Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

(C3) Full and Frank Disclosure

Oral and written evaluation reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.

(C4) Public's Right to Know

The formal parties to an evaluation should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.

(C5) Rights of Human Subjects

Evaluations should be designed and conducted so that the rights and welfare of the human subjects are respected and protected.

(C6) Human Interactions

Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation.

(C7) Balanced Reporting

The evaluation should be complete and fair in its presentation of strengths and weaknesses of the object under investigation, so that strengths can be built upon and problem areas addressed.

(C8) Fiscal Responsibility

The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible.

(D) Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its worth or ment. These standards are:

(D1) Object Identification

The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly identified.

(D2) Context Analysis

The context in which the program, project, or material exists should be examined in enough detail, so that its likely influences on the object can be identified.

(D3) Described Purposes and Procedures

The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

(D4) Defensible Information Sources
The sources of information should be described in enough detail, so that the adequacy of the information can be assessed.

(D5) Valid Measurement
The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid for the given use.

(D6) Reliable Measurement
The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information obtained is sufficiently reliable for the intended use.

(D7) Systematic Data Control
The data collected, processed, and reported in an evaluation should be reviewed and corrected, so that the results of the evaluation will not be flawed.

(D8) Analysis of Quantitative Information
Quantitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

(D9) Analysis of Qualitative Information
Qualitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

(D10) Justified Conclusions
The conclusions reached in an evaluation should be explicitly justified, so that the audiences can assess them.

(D11) Objective Reporting
The evaluation procedures should provide safeguards to protect the evaluation findings and reports against distortion by the personal feelings and biases of any party to the evaluation.

Weir and Roberts 1994 : 246-7

APPENDIX 3.1

Formats for programme evaluation by participants

A POSSIBLE EVALUATION FORM

I liked:

I didn't like:

I suggest:

GRAFFITI WALL

For this you will need a page of 'bricks' in a 'wall' so that trainees can write their comments, graffiti-style across them (Lavery, 1985).

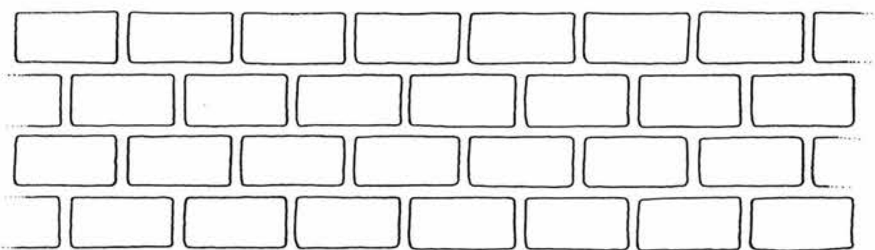


Figure 25 The writing on the wall

CONSECUTIVE SLIPS

People write three separate remarks about the course or an aspect of the course on three separate pieces of paper and put them in a box in the centre of the room. The papers are all stirred up and then people pick out three papers from the box. Make sure nobody has picked out their own paper again. Somebody starts by reading out the comment written on the piece of paper they have picked out. Other people listen. If someone listening feels that the comment is similar in topic and mood to any of those on their pieces of paper, then they read out their related comment.

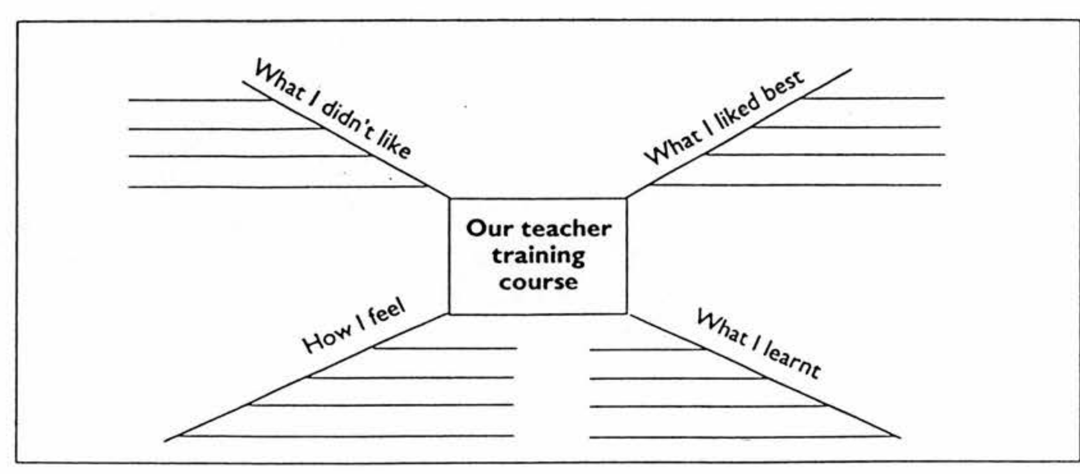
Thus, if the first comment which is read out relates to, say, the resource library, then other people will start to read out any comments they have that relate to the resource library or to resources in general. When that topic is exhausted, someone else reads out a fresh comment on a new topic. There may be a few comments left at the end that do not relate to anything anyone has so far read out. They are simply read out last[1].

Woodward 1991: 123, 126-7

Feedback sheet 2 (a)

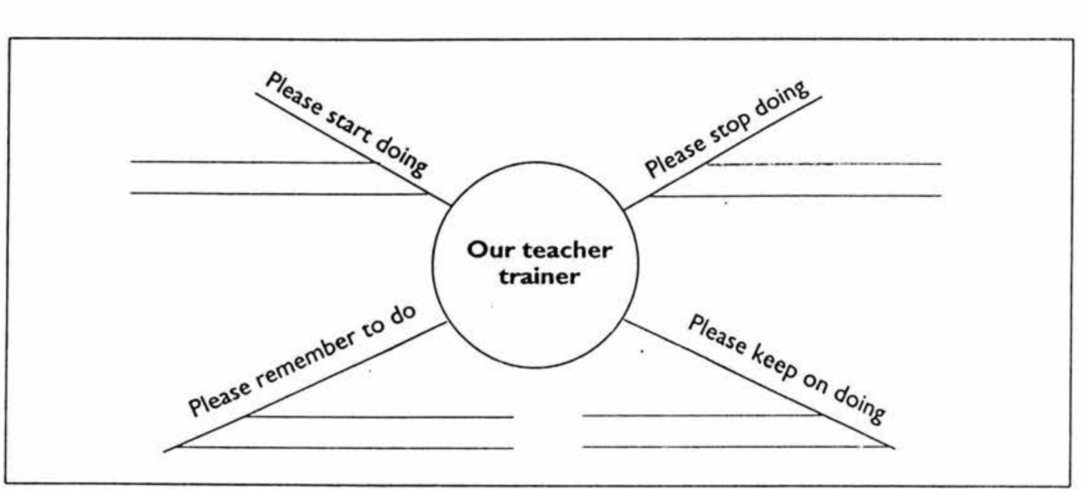
Complete the following mind maps

Feel free to write exactly what you think/feel



2 (b)

Add as many sub-branches as you need



Vasconcelos 1994: 13-14

APPENDIX 3.2

Example of Rapid Assessment Form

A = Much above average
B = Above average
C = Average
D = Below average
E = Much below average.

RAW / WEIGHTED

20 participants	A (+2)	B (+1)	C (0)	D (-1)	E (-2)	TOTAL	RANK
Organization: programme structure	9 / 18	10 / 10	1 / 0			28	=5
Organization: presentation	9 / 18	11 / 11				29	4
Content: interest	13 / 26	7 / 7				33	2
Content: relevance	8 / 16	12 / 12				28	=5
Quality of internal speakers	8 / 16	12 / 12				28	=5
Quality of external speakers	3 / 6	14 / 14	2 / 0			20	11
Use of audio-visual aids	3 / 6	12 / 12	5 / 0			18	12
Use of other resources	2 / 4	9 / 9	7 / 0	1 / -1		12	14
Documentation	9 / 18	8 / 8	2 / 0	1 / -1		25	=9
Participation	8 / 16	10 / 10	1 / 0	1 / -1		25	=9
Group Discussions	10 / 20	6 / 6	3 / 0			26	8
Stimulation of thought	16 / 32	4 / 4				36	1
Stimulation of self-examination	11 / 22	9 / 9				31	3
Displays		4 / 4	11 / 0	1 / -1		3	16
Creature comfort: Accommodation	2 / 4	7 / 7	8 / 0			11	15
Creature comfort: Meals and drinks	5 / 10	6 / 6	8 / 0			16	13

Rudduck 1981: 85

Appendix 3.3

Action letters

Letter to My Boss

Dear ____________________:

I have just completed a training program entitled "________________ ______________________________." I want to tell you what I feel I learned, and how I plan to change or improve as a result. I would appreciate talking to you about the following ideas in the near future. I will then solicit your active support in implementing these changes.

Here is what I propose to do differently/better:

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

Signed__________________________

Date____________________________

Newstrom and Scammell 1980: 299

Letter to the Trainer

Dear ____________________:

I attended the ____________________________________course conducted by you on _________________________, 19___. I wish to share with you a series of insights I have gained since then regarding the ways in which I have (have not) been able to apply the material to my job.

I have done the following:

1.

2.

3.

Information in these areas has proven to be of considerably less use to me:

1.

2.

3.

The suggestions I have for you include:

1.

2.

3.

Signed______________________________

Newstrom and Scammell 1980: 295

APPENDIX 3.4

Briefing notes for implementation plan

UNIVERSITY OF LANCASTER:
INSTITUTE FOR ENGLISH LANGUAGE EDUCATION,
ESPTT COURSE, OCTOBER–DECEMBER 1982

IMPLEMENTATION PLAN

Aim: To prepare a personal plan of action for further ESP curriculum development in your home institution (and beyond).

Rationale: During your time in Lancaster you have been asked to think about various aspects of the language learning and teaching curriculum – characteristics of the good language learner, syllabus design, textbook evaluation and test development to name but a few. We feel that you have learned a great deal on this course, and it is very important that you share this with others and put it into practice. For this reason we would like you to prepare a short report, to be handed in on Wednesday, 8 December, at 9.30 a.m. – with an outline and details of how you will apply what you have learned when you return to your home institution.

Your report should focus on those aspects of the curriculum that you feel you can most contribute to – any aspects that we have examined in class and others such as teacher training. If you feel that your ESP teaching programme is in need of improvement and that you have a part to play in this development it is important to capture your ideas *now*, while they are still fresh in your mind.

Procedure: What is important is to make the plan realistic in terms of what you feel can be expected to be achieved. It should not only be something that you need and want, but also something which is likely to be carried out. We intend to follow up your plan by contacting you within the next few months to see if you have been able to carry out your plan and with what success.

There will always be resistance and constraints, so in your plan try to outline what they might be and where they might arise, and what ways there may be for overcoming and accommodating them. Think hard in your plan how you can get round obstacles that might appear to prevent your plan being carried out.

You will need to examine carefully and state the resources needed to carry out your plan, both human and material.

In short, the main points to take into account are:

(i) how to get round constraints;
(ii) how to disseminate information;
(iii) how to persuade others involved;
(iv) how to acquire the necessary resources.

Your plan should include a *justification* of what you propose to do in terms of what you know about the nature of language learning and the curriculum.

Alderson 1985: 148-9

APPENDIX 3.5

Example of course log

Name:

Course Title: **Course Dates:**

This log serves a threefold purpose:

- It helps you to reflect on what you have learned and in what ways you have benefitted from each day of the course
- It provides you with a record of your personal progress which you can use for your course debrief
- It provides feedback for your tutors on the course and on their performance

The diary section is yours to keep.

The questionnaire will be collected at the end of the course.

Expectations
Write a brief outline of what you expect to gain from the course.
Include any points covered in your course briefing.
Be prepared to discuss your expectations with other members of the course.

Day 1

etc, depending on the number of days or sessions on the course.

Evaluation Look back to your *expectations*, read through your daily review and comment on what you have gained from the course as a whole. Be prepared to discuss your evaluation with other members of the course.

Buckley and Caple 1990: 237-8

Appendix 3.6

Student response slip

TESOL CENTRE

Student Response Slip

Subject/unit ______________________________ Module number ______

Please tick the boxes and/or give your views:

1) I understand your comments ☐

____

____

____

2) I agree with your comments ☐

____

____

____

3) I agree with the grade ☐

____

____

____

Name ______________________________ Course __________________

Parker and Graham 1995: 168

APPENDIX 4.1

Letter to institutions

INSTITUTE *for* APPLIED LANGUAGE STUDIES

The University of Edinburgh
21 Hill Place
Edinburgh EH8 9DP
Scotland

Telephone 031 650 6200
Fax 031 667 5927

March 29, 1993

Dear

I am conducting research into the means by which UK institutions **evaluate their courses for practising teachers of foreign languages** (EFL and Modern Languages) and would be grateful for information on the procedures used by your institution. My report will only make reference to specific institutions if permission has previously been obtained.

As you will see, there are two forms to complete: the green one simply requires you to list and classify the in-service courses you run; the white sheet (multiple copies) contains a few questions on how the courses are evaluated. If you run only one or two courses, it should take no more than 10 minutes to complete the forms.

It would be very helpful if you were able to respond **by the end of April**. However, if you run so many courses that a full response would be very onerous, perhaps you could simply return the Courses Checklist and I will give you a ring in May to get the relevant details.

Thanking you in anticipation,

Yours sincerely,

Ian McGrath

Development Coordinator
Teacher Education

APPENDIX 4.2

Courses checklist

NAME OF INSTITUTION .... TEL. No. ....

ADDRESS ....

....

CONTACT PERSON FOR INFO. ON TEACHERS' COURSES ....

Please list the different courses run (or to be run) by your institution in the year July 1992-June 1993 for practising teachers of EFL or other languages. Indicate by ticks in the appropriate column how each course can be classified.

KEY TO CLASSIFICATION

UK	based wholly in the UK
O(verseas)	based wholly overseas
UK + O	taught partly in the UK
UK + D(istance)	distance-learning course with UK component
D	wholly distance learning
Open	open enrolment
Con.	contract (closed) group
F-T	full-time (min. 3 hrs per day)
P-T	part-time
Q (E)	leading to external qualification (e.g. Cambridge/RSA Dip TEFLA)
Q (I)	qualification awarded by institution running course

NAME OF COURSE	UK	O	UK+O	UK+D	D	Open	Con	F-T	P-T	Q(E)	Q(I)
....											
....											
....											
....											
....											

[*space for further courses overleaf*]

Please return to: Ian McGrath, IALS, University of Edinburgh, 21 Hill Place, Edinburgh EH8 9DP (tel. 031 650 6200)

NAME OF COURSE	UK	O	UK+O	UK+D	D	Open	Con	F-T	P-T	Q(E)	Q(I)

APPENDIX 4.3

Questionnaire

EVALUATION OF COURSES FOR LANGUAGE TEACHERS

Please complete a separate form for each course described.

INSTITUTION ....

COURSE TITLE ....

COURSE ESTABLISHED (date) .... FREQUENCY (x per year) ....

AVERAGE NUMBER OF PARTICIPANTS ....

Please indicate with a tick (✓) which of the following procedures are used to evaluate this course.

E. validated by external examiner — E. [ ]
R. results (number and level of passes) used to evaluate success of course — R. [ ]

1. **pre-course analysis of existing knowledge, skills, wishes, etc** — 1. [ ]

by means of:
1.1 questionnaire — 1.1 [ ]
1.2 observation of participants' teaching — 1.2 [ ]
1.3 interview — 1.3 [ ]
1.4 test — 1.4 [ ]
1.5 information from sponsor — 1.5 [ ]
1.6 tutors' experience of previous similar groups — 1.6 [ ]
1.7 tutors' experience of country concerned — 1.7 [ ]
1.8 other (please specify) — 1.8 [ ]

2. **formal in-course monitoring** (i.e. during course) — 2. [ ]

by means of:
2.1 formal discussion with whole group at regular intervals — 2.1 [ ]
2.2 participant questionnaire — 2.2 [ ]
2.3 individual participant interviews to discuss progress, attitudes, etc — 2.3 [ ]
2.4 tests/exams/written assignments — 2.4 [ ]
2.5 observation of sessions (e.g. by Course Director) — 2.5 [ ]
2.6 meetings of course tutors to assess progress of course and individuals — 2.6 [ ]
2.7 regular written tutor reports — 2.7 [ ]
2.8 other (please specify) — 2.8 [ ]

p.t.o.

3. **end-of-course evaluation** 3. [ ]

by means of:

3.1 formal discussion with whole group 3.1 [ ]
3.2 participant questionnaire 3.2 [ ]
3.3 individual participant interviews 3.3 [ ]
3.4 tests/exams/written assignments 3.4 [ ]
3.5 meeting of tutors 3.5 [ ]
3.6 written tutor reports 3.6 [ ]
3.7 other (please specify) 3.7 [ ]

4. **post-course evaluation** (e.g. 3 months after course) 4. [ ]

by means of:

4.1 participant questionnaire 4.1 [ ]
4.2 report from sponsor or other responsible person 4.2 [ ]
4.3 observation of participants' teaching 4.3 [ ]
4.4 follow-up meeting for participants 4.4 [ ]
4.5 other (please specify) 4.5 [ ]

---

5. Are you satisfied, in general, with the procedures currently used to evaluate your courses for teachers ? YES 5.1 [ ] NO 5.2 [ ]

If NO, please explain why you are dissatisfied. 5.3

Thank you for your help.

Please return to: Ian McGrath, Development Coordinator: Teacher Education, IALS, University of Edinburgh, 21 Hill Place, Edinburgh EH9 1JG

APPENDIX 4.4

Summary of Evaluation Procedures Used, broken down by evaluation stage and course type

		M	B	Dip	OC	NC cl	NC op	TOTAL
		21	5	10	18	25	31	**110**
E	validated by external examiner	20	5	4	17	1	-	**47**
R	results (number and level of passes) used to evaluate success of course	7	4	10	11	2	1	**35**
1	**pre-course analysis of existing knowledge, skills, wishes, etc** by means of:							
1.1	questionnaire	-	2	3	5	20	23	**53 (48%)**
1.2	observation of participants' teaching	-	-	3	1	3	-	**7 (6%)**
1.3	interview	13	2	10	6	3	5	**39 (35%)**
1.4	test	12	3	3	7	3	10	**38 (35%)**
1.5	information from sponsor	14	4	2	11	17	4	**52 (47%)**
1.6	tutors' experience of previous similar groups	17	4	7	14	19	24	**85 (77%)**
1.7	tutors' experience of country concerned	12	5	1	5	12	10	**45 (41%)**
1.8	other							
	• references	6		1	2			**9**
	• letter of application	3						**3**
	• application form/work record	5			1			**6**
	• pre-course tasks as part of interview			1				**1**
	• meeting of CDs from other institutions					2		**2**
	• knowledge of market						1	**1**
	• visit to country (briefing/CD meeting/school visits/in situ needs analysis					4		**4**

		M	B	Dip	OC	NC cl	NC op	TOTAL
		21	5	10	18	25	31	**110**
2	**formal in-course monitoring (i.e. during course)** by means of:							
2.1	formal discussion with whole group at regular intervals	19	4	8	13	20	20	**84 (76%)**
2.2	participant questionnaire	15	4	6	13	19	17	**74 (67%)**
2.3	individual participant interviews to discuss progress, attitudes, etc	18	3	9	13	12	17	**72 (65%)**
2.4	tests/exams/written assignments	19	5	10	17	9	4	**64 (58%)**
2.5	observation of sessions (e.g. by CD)	4	2	5	8	12	16	**47 (43%)**
2.6	meetings of course tutors to assess progress of course and individuals	14	5	8	13	22	17	**79 (72%)**
2.7	regular written tutor reports	12	3	3	6	7	2	**33 (30%)**
2.8	other							
	• student representatives/staff-student committee	5		1	2			**8**
	• monitoring of tutorial work	1		1				**2**
	• external evaluation by staff from other parts of Faculty	1						**1**
	• visits by moderator					1		**1**
	• 6-monthly reports to sponsors		1					**1**
	• regular sponsor 'inspections'		1			1		**2**
	• tutor reports on participants' teaching			1				**1**
	• memo from CD to course team			1				
	• group leader observes and coordinates with CD					1		**1**
	• weekly staff meeting reports					3		**3**
	• poster presentations on projects					1		
	• diaries						3	**3**
	• regular meeting of CD and TT coordinator					2		**2**

		M	B	Dip	OC	NC cl	NC op	TOTAL
		21	5	10	18	25	31	**110**
3	**end-of-course evaluation** by means of:							
3.1	formal discussion with whole group	12	4	8	11	21	22	**78 (71%)**
3.2	participant questionnaire	18	5	6	16	23	25	**93 (85%)**
3.3	individual participant interviews	12	3	2	5	7	5	**34 (31%)**
3.4	tests/exams/written assignments	16	5	6	8	5	-	**40 (36%)**
3.5	meeting of tutors	14	5	8	13	21	17	**78 (71%)**
3.6	written tutor reports	4	3	6	5	18	6	**54 (49%)**
3.7	other							
	• dissertation	1						**1**
	• CD's report	1		2	2	1	2	**8**
	• debriefing of CD by TT coordinator					7	4	**11**
	• optional meeting with external examiner				1			**1**
	• exam board	5	1		2			**8**
	• external examiner's report	4	1		4	1		**10**
	• report to sponsors		1	1	1			**3**
	• formulation of action plan by course team	1		1	1			**3**

		M	B	Dip	OC	NC cl	NC op	TOTAL
		21	5	10	18	25	31	**110**
4	**post-course evaluation** by means of:							
4.1	participant questionnaire	2	-	1	1	5	3	**12 (11%)**
4.2	report from sponsor or other responsible person	1	2	-	1	8	-	**12 (11%)**
4.3	observation of participants' teaching	-	1	1	-	4	-	**6 (5%)**
4.4	follow-up meeting for participants	-	-	1	-	7	-	**8 (7%)**
4.5	other							
	• letter to participants					1		**1**
	• external examiner's report	4			1			**5**
	• follow-up visit by CD /tutor to monitor training/assess project/observe teaching					4		**4**
	• UK course part of ongoing project, which is monitored					1		**1**
	• personal contact with individual students				1		1	**2**
	• letters from participants						1	**1**
	• participants return for another course						2	**2**
	• recommendations to other teachers						4	**4**

APPENDIX 4.5

Level of satisfaction with evaluation procedures used

[M: Master's; OC: other certificated; NC cl: non-certificated closed; NC op: non-certificated open]

UNIVERSITIES							
	M	B.Ed	Dip	OC	NC cl	NC op	TOTAL
YES	6	1	2	5	2	4	18
YES/NO	3	1	1	5	-	1	11
NO	8	2	-	3	-	-	13
n.r.	1	-	-	1	-	2	4
TOTAL	18	4	3	14	2	7	46

COLLEGES							
	M	B.Ed	Dip	OC	NC cl	NC op	TOTAL
YES	-	-	1	1	2	3	7
YES/NO	1	-	-	1	-	-	2
NO	2	1	2	-	-	-	5
n.r.	-	-	-	-	3	-	3
TOTAL	3	1	3	2	5	3	17

LANGUAGE SCHOOLS							
	M	B.Ed	Dip	OC	NC cl	NC op	TOTAL
YES	-	-	3	2	7	14	26
YES/NO	-	-	-	-	1	1	2
NO	-	-	-	-	-	2	2
n.r.	-	-	1	-	10	14	25
TOTAL	-	-	4	2	18	31	55

APPENDIX 4.6

Interviews with Informants 1-5

Edited transcripts

INFORMANT 1

What do you see as the purpose of the evaluation that you do actually carry out ?

1. I can immediately I can think of three reasons [*Hm*] not in order of priority. One is quality assurance [*Hm hm*] one is for us to ensure that we are providing the goods that our clients require and I suppose that we are seen to be doing so and there's a sense also which it's if it were needed as a motivating factor for the staff involved, the fact that there is there is some kind of there is some kind of evaluation of courses which is external to the staff involved in the course. [*Negative motivation ?*] No I don't think so. That's one reason. Another reason ... and this is summative, end of course evaluation, [*Hm*] changes which which we make to the organisation or content of courses erm often is a response to themes which have come up again and again, quite specific things [like changing the time of a course from p.m. to a.m.]. The third reason and the third reason is developmental observation particularly on the longer courses we change erm the content and the way we run the courses according to feedback we get in evaluating them at all stages of the course. [*Hm*] Erm and quite a lot of the courses we run particularly the open the open-access ones er begin with an analysis of needs and then there is a process of erm fine-tuning and modifying that analysis of needs all the way through the course erm so that the course participants are in a sense responsible for the content of the course.

There are two questions arising from what you've just said: erm one has to do with the with the needs analysis which you mentioned in the forms the other has to do with erm the amount of information you you require to persuade you to make changes [Hm] *of a particular kind.* [Hm] *Would you would you make changes on the basis of a single run of the course?*

2. For subsequent courses ?

Hm.

3. It would depend. It would depend what it was. And it would depend whether we thought that the profile of the people on the course was was going to be er similar in a in a subsequent course. When we are running a course for the first time probably 'Yes' [*Hm*], where we have run a similar course over a period of time and then suddenly we get a particular erm point of view 'No' but but I mean I would er keep a record of that [*Hm*] and I would ensure that erm either that it came up explicitly as something we we asked people to address participants to address in a subsequent course or that the feedback on the subsequent course was compared [*Hm hm*] and that was looked at.

That implies record keeping [Yes] *of a particular kind* [Yes]. *Can you tell me a little about what sorts of records that you would typically expect to be kept ?* [Yes].

4. Erm at the end well at the end of every course somebody depending on the course who often me somebody will spend an hour with the course

participants erm doing what is disparagingly called a happiness sheet, I think, [*Hm ?*] although it's quite interesting that that erm if you're not looking for you can make them whatever you want. They're only happiness sheets if you're trying to get a record of happiness from them. So I don't think the ones we get are. Erm but somebody goes and gets feedback from them. They are they know this is going to happen. In that final week they're asked to reflect on certain criteria and they know that they will have the opportunity to give feedback. And typically they will in groups discuss the criteria and appoint a spokesperson [*Hm*] so when for example I go into the room they're primed and ready to speak and I see my role as being erm to say things like 'How many people agree with that ?'. And and I write notes as I listen to them [*Hm*] and then I read the notes back to them and they say "No actually it's not 'very' it's only 'fairly'". [*Hm hm*] That gets typed up and that record is kept. I would put if there have been problems during the course as there sometimes are there would be I would add a record of those problems to that written record. And I would say erm during the course the the group argued erm at several points that they wanted to be treated not as a single group but to be mixed with other groups and and this is not reflected in the final feedback but they felt it very strongly [*Hm*] during it.

How would you have known about that ?

5. Through the feedback that the course tutors hold of a similar kind throughout the course and through diaries. [*Through participant diaries ?*] Participant diaries. *Right, that's something I was going to come back to later as a particular instrument.*

Is it important within the systems you have here that happiness sheet evaluation is carried out by someone external to the course team ?

6. Yes. [*So that's the procedure.*] That's personal. That that is the policy but erm erm it's me who imposes it. [*Right*] Not everybody likes that. [*Hm hm*] ...

What about the criteria ? Are those suggested by you or the course team or are they criteria which the group discussions throw up ?

7. It depends it depends on the group and I suppose in a sense on their sophistication. One of the courses that we run as an open-access course is a course in teacher training skills and although the content of that is negotiated it's almost certain that there will be something during the course on evaluation of courses [*Hm*] and typically they will as a project they will devise the way they want their own course [*Right*] to be evaluated. In that in that case erm er there would be no I mean it would be entirely them. They would have decided how to do it. That's one end of the spectrum. At another end of the spectrum erm will be a group of people who for example didn't take responsibility and were not sophisticated or didn't work together well erm I might go in I might give them a very specific list of criteria. [*Hm*] I sometimes do very I'm sorry it's not terribly systematic. I sometimes what I tend to do with some groups and different people I'm sure do it differently one thing I'll do - I'm thinking of 2-week 30-hour open-access courses - I'll go in and I'll say erm I'll state what the aim what the objective of the course was and what we hope that

they will feel that they've achieved and then say absolutely nothing [*Hm*] and then there's quite a long silence [*Hm*] and then somebody will say something ... and I will play it entirely by ear the only thing that I will really do is to ensure that everybody has the opportunity to say something.... On courses of 30 hours that would only happen at the end of the course or if a problem arose during it.... On longer courses erm what I'm describing might happen more frequently. [Goes on to describe very problematic closed course where group's agenda and sponsor's were completely different. For second course with same type of group, regular 'listening' (to participants) sessions were introduced involving a course-external listener. Tutors had asked for this to happen.] They felt that they were getting feedback that they which required action that they were not in a position to take and they felt that if they dealt with some of the hostility in within the group it would compromise their relationship [*Hm*] with the group so they wanted somebody else to do that.

Going back to the needs analysis which is something you just touched on. Erm clearly in some cases cases of contract groups you you try to carry out an in situ analysis you mentioned here. Would that be standard, part of the costing and so on ?

8. Yes. *In situ* you mean *in situ* here or *in situ* there ?

There.

9. We've never we've we've never except [in a specific case] been able to do it in any form other than through qu-aire. And and written assignment. We've never been in the fortunate position of being able to send someone to observe....

Would you see that (in-situ analysis) as desirable ? What difference do you think it would make ?

10. Erm well in the case of [specific group] ... the objective was selection... People applied to be a part of a programme... The selection process wasn't to find out if people would be good enough. It was to find out whether the work they'd be doing when they returned would enable them to show that the skills they were acquiring during the programme had been of value...

Is it the case that in general you tend to operate fairly strict selection procedures and that the information that you gather during that process feeds in to a needs analysis which informs the course in some way ?

11. Yes. Yes for the open-access courses the application procedure involves them writing an essay on why they want to do the course. [*Um*] Erm the tutors will use that to draw up the provisional course programme for the first part of the course. Erm and on typically on the first day of the course there will be a more detailed analysis and discussion of needs and interests. And they will be presented with that programme for the first part of the course and asked to comment on it in some way. And that typically will then be modified. [*Um*] And then subsequent stages of the course will be determined by course participants... [Explains that 'this can be a bit of a con' in that participants can be persuaded to vote for

things they'd never thought of, such as classroom research, which tutors know they'll enjoy and find useful] There are all sorts of problems with needs analysis. I mean the first problem is people not knowing their needs [Hm] and the second problem is people oh (deep sigh) the second problem is people using terms to mean completely different things [Hm] from what you use them to mean and the third thing is people saying what they think is expected of them to say and the fourth problem is erm people's needs bearing no relation to what people's wants bearing no relation to their needs and I mean I think it has to be a process of of negotiation.

Don't you think that the problems you've been describing there and the issue of honesty on the part of participants also figure as far as the end of course evaluation is concerned, whether it be a questionnaire or a or a discussion ? Isn't there a tendency to say what you think you're expected to say ? I should rephrase that because it sounds as if I'm putting words in your mouth. [both laugh]

12. You probably are because my instinct is to say 'No'. [laughs]

Do you you feel you can trust what people say to you - or write for you ?

13. I think it usually reflects what they feel then. [*Hm*] And the ideal - and in some cases we're able to do this but rarely - the ideal is to do feedback much later. [*Hm*] And then I think there's more honesty. But I think where I think there's a problem of people saying what what they think is expected of them is actually on courses like a CTEFLA course where there is an element of assessment [*Hm*] and they feel that if there is any doubt over their grade they may be prejudiced by criticising [Hm] the course... But I think I think perhaps this is utterly naive ... I think that the climate of openness and honesty and trust is such I ... I believe this ... in this place that by and large people feel comfortable about giving honest feedback.

What about this delayed post-course feedback ? What sort of feedback would you see as being the best ... ?

14. The best we were able to do this once ... is to visit people where they work and to see what they're doing and to talk to them. [Describes the background to a specific course] That was very useful, very encouraging.

What what would she have been evaluating ?

15. What she was specifically asking them was whether there were things they thought they could have done while they were here that would have been helpful to them that we didn't do. What kind of instruments she used can't remember. I suspect nothing very formal.

Can I ask you now about some of the in-course monitoring instruments. Erm diaries figure several times in the return. Can you tell me a little about how that's used ?

16. Yes. Varies from course to course and it varies from tutor to tutor.... Different tutors use them in different ways from those who who leave ten minutes for the at the end of each day for participants of silence while

participants write in the notebooks and the advantage of that is the great one of that is that the par the notebooks never leave the building [*They're collected in ?*] Hm ? [*They're collected in.*] They're collected in each day and the tutor will read them each day and write something in them each day and give them back. Erm two, two participants take them home and on Friday or at the week-end I think whenever they kind of finish it off and they hand it in and the tutor has one night and goes through them and gives them back. Erm some tutors give very detailed guidelines to the participants about what they feel they should write about, other tutors erm give no guidance at all and er some negotiate with participants about whether they correct mistakes what kind of comments they make some some kind of [?] absolutist that this is [a kind of?] process and that they should only respond to the content of what people say erm some tutors negotiate with participants at the beginning of the course whether or not they want to use diaries. And sometimes it's awful. A course a colleague did in September which was a negotiated content course on Thursday evening of the second week - the course finished on the Friday - the Thursday evening of the second week two people who were friends and had come together ... wrote in their diaries that they hadn't got what they had wanted out of the course and she was really distressed and upset and hurt. This rapidly changed to to anger I mean she was furious because they had set it up so that their needs wouldn't be met. [*Hm*] That's what people do in relationships. [...] And [when she saw them again] she didn't respond honestly. She didn't say "I'm furious that you've done this". She said "I'm sorry that you felt that your needs weren't met". (laughs)

What do you feel comes out of the diaries in evaluation terms ?

17. You learn people say what they like and what they don't like and what they feel they need. People often say very interesting things like in groups where there are problems within the group they write about the problems within the group and their problems with each other - that's not really what you want, no. Erm depends how it's depends on the group they people some some people write some people write to please and they they just say everything's wonderful and what they like it's not terribly useful but I don't think there's much you can do about that erm

This is confidential information in a sense but presumably from time to time tutors may feel the need to act on it rather than simply replying in writing [Hm]. *Is there a an ethical problem there about surfacing in the group something which an individual has in a sense said to you privately ?*

18. Depends on what the contract is with the group over the diaries. I don't imagine a tutor would do that without speaking to the person and getting their permission [*Hm*] ... We had a closed group earlier this summer ... their diaries were vitriolic, about each other, to the tutors about the tutors and about the course. And basically they felt that they had been ... trickedin attending a course that they hadn't wanted to attend and that they felt that ... they had been misled as to the objectives and content of the course, deliberately in [COUNTRY]... There wasn't much we could do in response except listen and counsel ... You can't always respond, you can't always do anything at all, but it's much better to know what it is that they're angry and unhappy about than just to kind of see an angry and unhappy group of people.

Is the unhappiness sometimes about being more or less obliged to write a diary ?

19. No. [*No ?*] Because they don't have to do it if they don't want to.

But doesn't that then undermine the effort as far as the rest of the group is concerned or the efforts of individuals within the group ?

20. Why ?

Some people may accept it as being one of the house rules without wanting to do it but then feel that well I'm beginning to see a value in this and actually appreciate it. Whereas if individuals within the group were being allowed not to do it you know they might the haverers the waverers might join them.

21. Hm. Yes I've never really thought about that.

But it hasn't been a problem ?

22. In in courses that I have run I've allowed people not to write their diaries [*Hm hm*] And I've always kind of thought that you know that if they don't if they have something to say there is a channel they know there have a channel in which they can kind of say it and address it to me. If they choose not to that's fine.... A journal I call it but a lot of my colleagues call them erm "thinking books". [The interviewer then shares an experience of asking learners to keep diaries on a course designed to encourage autonomous learning] I think I think there's an important distinction between self-direction and the diary as a contribution to that and a course where people have been consulted. The programme is is centrally determined and the one of the functions of the diary is to to keep a check on how people are reacting to that [Hm hm]. And here that is more the case ... there is a sense in which you spot the individuals who who are dissatisfied ... in any group there's going to be one or two people whose needs can't be met in the group you can erm "massage" them and give them books to read [Hm] on topics that they want and arrange for them to go and talk to your colleagues who're particularly interested [Yes] in the topic. No that's not massaging them that is meeting their needs [Hm] to some extent.

I don't know whether it's possible to generalise here, but can you say which of the evaluation measures you use is most productive ? [In terms of?] *Gathering information for whichever of the initial purposes you mentioned.*

23. Well, for the quality control ... the summative evaluation at the end when another person is there with the group; for erm making changes to subsequent courses post-course feedback ... and for making changes during the course probably diaries mediated with discussion which arises between the tutor and the people which arises taking into account what she or he has read in the diaries. Another thing we do is we have a competition and we give a prize and the prize is a free course [*and the competition ?*] ... the competition is to suggest ways in which we can improve the courses we offer [Hm] and we get we get lots of feedback.

Have you modified the evaluation procedures you use over time ?

24. Yes. You're going to ask me how aren't you ? [*How and why I was going to ask.*] Erm on the special group courses there is more evaluation and it takes place it takes place erm at frequent intervals during the course and that involves discussion between tutors and the course participants like each Friday there's a discussion of what people have done what they feel they've learnt [what or who helped them?] and what they'd like to do and that is because of my kind of horror of discovering at the end of a course that people feel their needs haven't been met ...

One of the things I read suggested that as far as refresher courses are concerned and a lot of general summer courses for teachers would fall into that category the main criterion for evaluation should be whether or not teachers feel refreshed. How would you respond to that ? Do you feel it is enough to assess that ?

25. I think that's quite important (pause) No I mean the nicest thing is ... you go in and say the aim of this course was that after these two weeks you will feel you have some ideas you can implement in your teaching that you feel more enthusiastic about your teaching you feel more confident about your teaching [*Hm*] that you feel that what you do has been confirmed er silence and somebody says "I feel all those things". [*Hm*] That's so good ... what they're saying really is "We feel refreshed, we're going back with" (tails off) There's a big difference between people who choose to come on courses and people who've (tails off)

These are all attitudinal things, aren't they ? [Yes] *They're not observable changes in knowledge, understanding, skill, etc. Do you feel at all uneasy ?*

26. As you say that, terribly ! ... Yes closed groups who come here usually work on projects and they actually work on a project that they can take home with them and be bound together in a book [Hm] that they'll each have a copy of each person's project [*Hm*] . I don't actually think of that as part of the evaluation. That's very concrete and observable and it's it's what they're putting together is evidenve of something they have done that they couldn't have done before.

I have a feeling that that way of thinking also may pervade the way we think about long courses as well [Hm] *except for things like er Dip TEFLA which are assessed by other means anyway.*

27. Hm yes we could be very specific and ask people what they feel they've learnt.

So just trying to summarise is it the case that what you're really interested in is whether participants were satisfied with the course they had ? Is that is that really the the sort of centre of all the evaluation irrespective of whether it's directed to course development or ... ?

28. No it's it's not central to course development. It's not central to the evaluation which takes place during courses. It's central it's it's a key factor in what we do at the end of a course.

Hm Why is it not central to course development ? ... Wouldn't it be dissatisfaction which would prompt you to make changes ?

29. Well yes, but during the course they can say ... we would like to do something on this [*Hm*] or we would like a methodology which is much more directive from you and I don't think of that as satisfaction/ dissatisfaction I suppose it is ... during the course I'm not looking for them to say we feel we got our money's worth or we feel that the time we're spending here is is valuable and worthwhile. As I say that I realise that perhaps I should be. At the end of the course at the end of the course I really want the reassurance that they feel the course has been worthwhile and they have gained a lot from it [*Hm*] and it's important [...] to know the extent to which they have felt that and if they haven't to know why and what we could have done differently ... I'm realising how little I've thought thought it through as I talk (pause) and (pause) I think that if at the end of a course a group is really happy I kind of think it doesn't really matter what they say or what the details are we've done our job and that's it and I don't really find out what it was that made them happy [...] if they've been dissatisfied then I'm much more concerned.

But in saying that are are you saying you don't have quantitative data on which to make decisions ? Because you do end-of-course questionnaires which would supplement the data that you get from the discussion.

30. Impossible to process.

Impossible did you say ?

31. Yeah. The way they're organised, yes.

Are they open-ended ?

32. Hm. Totally ! [shows examples]

Would you feel I mean are you saying then that if it's difficult to process these results in quantitative terms you're less than happy with this format? Does it suit your purposes ?

33. I don't know what the answer is. Erm. Part of me is sitting here feeling a bit embarrassed and feeling feeling I think I've been rather smug about what we do. Part of me is sitting here and thinking (pause) I mean I haven't really thought through the question so I don't have an answer to it. My instinct part of my instinct is to say that er quantitative data (pause) is not so useful (pause) but maybe it is maybe we should [produces quantitative data for EFL classes showing level of satisfaction with various elements of course] ... we actually give them a questionnaire and we actually work it out in percentages ... we actually compare from month to month ... we don't do anything like that for teacher training (pause) No I mean I'm sorry I feel embarrassed [*No no you shouldn't*] I haven't really I haven't I've I've been fairly kind of satisfied by the way we do feedback without really or the way we evaluate courses without really thinking about it my instinct my instinct you see and this is the kind of smug side of course my instinct is that that we are (pause) that we are in touch with how people feel about their course [*Hm*] of course it doesn't that's

not the same as being in touch with what they're learning during the course and I think that instinct is right.

That then goes back to what objectives have been agreed, negotiated, etc with them [Hm] *and if you feel and they feel that those objectives have been achieved then in a sense what you haven't done is is perhaps irrelevant to what the course was all about.*

34. We don't actually we don't actually find out in any very systematic way what those objectives were. [Shows examples of feedback sheets completed by participants and seen by tutor who goes in to do final evaluation.]

There's no attempt to kind of reduce these to a coding frame and tally them - so many people mentioned this ? [No] *I'm not saying there should be.* (laughs)

35. No I know you're not no I know you're not I know you're not but you're you're you're kind of you're on virgin territory some of these thoughts haven't entered my brain before.

36. [Interviewer thanks informant, who suggests he might do talk for IATEFL Teacher Training Special Interest Group] ... the fact that I'm supposed to be fairly experienced in teacher training and I'm sitting here I'm thinking "Oh God" I think is indicative that this is probably an area in which quite a lot of people would have a lot to learn.

INFORMANT 2

[Types of course run - and evaluation procedures used ?]

1. We have er everything from er short two - by short I mean two weeks is a minimum - er short courses right the way through three-month courses to erm one-year courses and erm that would be the Master's or whatever ...

Do you consider the Master's programme to be INSET in the normal sense of the term ?

2. Yes, there's a sense in which all of the Master's students that we get are pretty experienced and so they er it's a kind of INSET certainly, professional development. I I think there's a kind of view of INSET as something that er you go and do on a short-term basis and er yet all of our Master's programmes are for experienced teachers who're involved in erm professionally upgrading themselves in some way. And the evaluation procedures vary. I mean there is always an end-of-course questionnaire. And there is during the course constant erm constant talk and erm even er on the Master's course once a term er a formalised an organised meeting where people get together - the course tutors and the participants get together - and er discuss the the way the course is going.

All the participants or just a committee?

3. Yes, all. It's called a course committee meeting but erm the course committee consists of all participants on the Master's course plus all tutors on the Master's course.

Does that is that a deliberate policy to include everybody ?

4. Yes.

Ah ha. Have you tried other approaches such as representatives ?

5. Yeah ... [but] on the M.Ed. course we've stuck with erm involving everybody.

And there hasn't been any problem about individuals feeling inhibited about speaking out in front of the tutors ?

6. Yes, I think I think that there have been problems with individuals speaking in front of the tutors. That comes out in tutorials. Every tutorial that we hold during a Master's course is an informal evaluation. You always get feedback.

Are you talking about individual tutorials [Yeah] or groups ? An informal evaluation because the tutor steers it in that direction ?

7. Not really, no. It comes. Erm, people always have things to say about the course sooner or later. You can have a tutorial where everything is er just dictated by say a student's need to make progress on on the dissertation, for example, but erm the way it comes out is unexpected

sometimes but in an individual tutorial what you very often get is erm somebody failed some kind of assignment and they feel sore about their grade or they want to understand the comments that the tutor 's made and erm the tutor then is pushed by the student to erm to give more information about about the grade and about the criteria he used to evaluate the assignment. And that is then seen in relation to the influence of the course and er to the student's individual perceptions of what the course is about. So we find ourselves actually in constant dialogue with our Master's students. Another thing we do on the Master's courses is to have an er an evaluation six months later. We send out a questionnaire to all of our Master's students in March each year erm to the last year's Master's group when they're back in post and we get very valuable information about what the course has meant to them. Erm once they're back in post. The reason why that why that is is simply because erm they er often can't get the course in perspective for themselves until they've returned to their posts and started to think about what the course meant for them.

Can you put a figure on how many the percentage of students that actually respond to this questionnaire ?

8. 95.

Really ? As many as that ? Do you have any particular technique for getting this very high response rate ?

9. I think it's something to do with the relationship we establish with them when they're with us ... We've never yet had a situation where erm people haven't wanted to write back to us ... They're glad to hear from us, they're glad to know we're still taking an interest in what they're doing ...

Is the questionnaire packaged with information about what you're doing and general information ?

10. We send a letter with the questionnaire. And the letter is a chatty sort of a letter which says things about what's going on in in er [INSTITUTION] and in er and how the present course is going - because by the time we send the questionnaire out erm we're always well into the next course. And erm then there's usually a handwritten bit at the end of the letter which is to do with them because we get very close to them during a course and er the handwritten bit is to do with the things that we know about about them and their worries or their expectations when they go back to their posts.

Is this something that you worked your way towards over a period of time or has this approach always taken this form ?

11. It's taken this form ever since we started the teacher training Master's because we we were very clear right from the very start of that course that it would have been premature to evaluate the course on the last day and so we we've always felt that erm our students wouldn't fully understand what the course meant to them until they got back to their own countries and started doing the things that were expected of them when they returned to post and I think that's been borne out quite often. And the

emotional involvement of the student in a course during a course is so high that you get some very distorted data from mid-course and end-of-course evaluation procedures, which have to do with personal agendas... If a student feels sore about a grade nothing looms larger than that at that moment. Just to give you a an example. We had a student from [PLACE] who did really well on the course but had a couple of duff grades on the assignments and she was very sore about those grades and there was some grief - I mean we had hard tutorials with her - and we had a letter back from her with a six-month-on evaluation form simply to say "Everything's in perspective now. I'm back in my post. I actually realise now that erm the grades didn't mean very much erm the experience meant everything and what I'm now doing - she's the in-service coordinator for all English teachers in [PLACE] - what I'm now doing is far more important than all of that. And I couldn't have done what I'm now doing without the course that I got.... Having to kind of sum that up, these six-month-on evaluations mean a lot more to us erm particularly because a lot of people are in projects. They have key posts in projects... what I personally feel about evaluation of the kind of course that we run - because they're very process-oriented - is that erm it takes a long time for them to think over what they've been involved in. Erm even six months might be too early for some of them.

I was going to ask 'Why why six months ?'. Is there not a danger that I mean clearly this is not borne out by your experience but theoretically after six months the ties with the institution are starting to weaken [Yes] *and after three months you know that would be less true* [Yeah]. *Do you did you so you didn't actually experiment* [No] *with the timings for these questionnaires ?*

12. We haven't done. I mean six months was just was chosen really for just that very purpose. We felt that the ties with the institution would have weakened and erm that people would not be feeling nostalgic about being in [PLACE] and everything else and they would have come to terms with the reality that they operate in....

So you were trying to counteract a nostalgia for [PLACE] [Yes] *by delaying ... ?*

13. And also trying to counteract the erm the feelings that the Master's course was the most important thing that they'd ever done. In a sense I mean it is the most important thing for all of them but erm it has to be placed in in perspective. And when you're on a course living a course and it has a very high profile in your life you don't think about much else and when you go back and you start doing the things that you know you have to do for the rest of your professional career or for the foreseeable future erm the Master's course either helped or didn't help and that seems to be a very good time to act [Yeah]. There's enough loyalty to us to ensure that the questionnaires get answered but enough detachment to ensure that there's a lot of objectivity. So it seems to work. And we get good data back.

What does the questionnaire cover ?

14. Erm the types of things that we're obviously looking at are a little bit different from the things we look at at the end of the course or during

the course. We want to know what the relevance of the course is to the professional needs of the individuals so we will ask for example 'In what ways did the course help you to do the job you're now doing ?' We also ask things like erm 'Are there areas that we didn't cover in the course that would have been useful to you, knowing what you now know ?' Things like that. [*Hm*] And it gives a slightly different perspective on course evaluation.

Do you feel that this is your most important way of getting data on student satisfaction [On Master's courses, yes] *or whatever ?*

15. We don't do the same kind of follow-up on three-week courses that we have or three-month courses that we have. Erm on those courses we do an end-of-course questionnaire and we take that at face value.

Does that mean you're satisfied with it ?

16. No. What it means is that there's a limit to the kind of endeavour that we put into follow-up and evaluation and when we have a three-month programme of one sort or another we we recognise that people are just just about with us for those three months but they're still very much focussed on what they've got to do when they go back and erm that imposes limitations, say. We get the data that we can then there and then [*Hm*] I I think that if we were honest about I mean the problem is erm always with evaluation erm the more sophisticated you make your evaluation instrument the more demands it places on you as a lecturer as a team of lecturers to to do something about it. And erm there's just simply a limit to time and er a limit to er how many different areas of focus you can cope with at any one time and so on.

As a self-financing institution is your main concern in evaluation to establish whether participants are satisfied with - is participant satisfaction the main criterion ?

17. It's one criterion but there is also I suppose I'm going to elaborate on what I said before there's also the fact that we deal with sponsors and those might be either erm ministries or erm British Council overseas offices or other bodies and erm when we when we look at what we've done with a particular group we're very interested to hear from the sponsors about the effectiveness of the course.... Where that becomnes particularly crucial for us is where we hope to do repeat business. And there've been over the last few years a number of occasions where we've run a course and have wanted to do a repeat course because we've known that the finance is there or the project has a certain life and it's very important for us to get it right next time round if we didn't get it right first time round. There is in fact a kind of negotiation between ourselves and the sponsors to improve in certain aspects of the course or change certain aspects of the course ....

It sounds to me from what you've said that sponsors are relying - as you rely - on participants' perceptions of the value of what they've done [Yes] rather than any let's say changes in behaviour in classrooms, which is presumably what your're really part of your ultimate aim is to change the

way people think about what they're doing and therefore change what they're doing and the way in which they do it.

18. I think you have to bear in mind erm that we work more with trainers than with teachers and erm that the er that the kind of measure of change, the behaviour, has more to do with trainer change than with teacher change. Of course, there's a knock-on effect and a common scenario for us is that we have to erm through the course for example this sounds pompous but turn teachers into trainers turn teachers into resource managers and what that means is that the criteria for evaluating what they're doing are established by the extent to which er the sponsor perceives that those goals those objectives have been met. Erm they'll send a teacher in to be a resource manager in [PLACE] for example. That teacher has no experience no prior experience of resource centre management and so after some time on the job they send them to us and we run a course which is focussed around resource centre management. And then they go back and erm we hear in all sorts of ways informal and formal how those people are doing in their jobs [*Hm*]. We try to establish quite close working relationships with particular countries which means that our network of contacts is quite refined and it also means that we get evaluation both formally and informally. So there'll be a formal route which is used when we write to the Ministry and an informal route which has to do with who you know out there. Erm so if somebody says it's made a difference [*Yeah*].

You mentioned the knock-on effect.... You're dealing mainly with trainers who deal with teachers who in turn deal with learners. Isn't this one of the problems for people who're trying to evaluate effects, that it's terribly difficult to get to the the sort of ultimate effects of seeing changes in learners as a result of what [Very very difficult] *has been done with teachers or trainers ?*

19. Very difficult because I mean six months would be much too early along the line to evaluate that and the knock-on effect er might take years [*Hm*].

Can we go to the other end of the spectrum, the the two-week courses that you run ? [Yeah.] *Do you evaluate these courses at all ?*

20. Yes, Just about. An end-of-course questionnaire. Erm the other way that we do it is that I suppose there's a lot of informal evaluation going on all the time with the participants on the course in that we have a pretty open-house policy at [INSTITUTION] where participants are encouraged to keep talking to the tutors about what's going on on the course erm and to give them feedback and at the end of a week in a two-week course we would sit round in a circle on a Friday afternoon ...

That's in addition to a questionnaire ?

21. At the end of the course. We wouldn't questionnaire them before the end of the course.

Have they filled in a questionnaire before the oral feedback ?

22. Well, the oral feedback erm would come at the end of the first week and then at the end of the course and usually at the end of the course erm

especially in a short course we'd ask for oral feedback before they fill the questionnaire in.

Is there a reason for asking for that bef...?

23. No, that's just the way it breaks.

Is that are the oral sessions conducted by the course tutors or by somebody outside the course ?

24. Usually it's done by the course tutor erm but the course tutor is one of a team of four or five people who've taught on the course. Erm if the course tutor has any reason to erm to suspect that things haven't gone that well on a particular course then he or she will call in somebody else to help with the evaluation. That has happened. [*Hm, but it's not a policy.*] No.

Do you have any views on this ?

25. No. I think it's a question of trusting colleagues really. People will have a duff course. We do sometimes. And when that happens erm people ask for help. People want their own feedback and they want to erm incorporate it into a course report and they want to be able to take the measure of it. That seems OK, that seems to work OK. Er so I think I feel fairly happy with that type of arrangement. [*Hm*] Everybody knows where they could go if they wanted to sound off about a course. And it does happen from time to time. When it happens erm there's a good network to take care of it.

We've talked about erm feedback from participants, feedback from sponsors via participants, erm you've just mentioned course reports by tutors. Erm are there any other sort of dimensions to the evaluation processes ?

26. Yes, there are. We're subject to external criteria as well and erm when when we run a course for a particular sponsor - I'll give you an example. We're running a course this summer for [NATIONALITY] teachers of English erm who're going to hold key training roles when they go back to [COUNTRY] but who up to now haven't held those roles and we're going to be visited by the erm ... the English Adviser to the project so he'll come in erm mid course - it's a three-month course so after six weeks - and will look at what we're doing and will discuss things with the students themselves and with the erm with the tutors on the course. And that's fairly common erm even during the course that er representatives of the sponsors will call in and actually spend time with the students, with the tutors and evaluate what's going on. It's quite important to do that because er well sometimes the students come from a culture er where deference to the teacher is all and they would say things erm behind the teacher's the tutor's back that they wouldn't say directly to the tutor's face [*Hm*] for fear of giving offence [*Hm*]. And we've seen we've on a number of occasions had valuable feedback from sponsors' visits visits during the course. Erm we also then get er valuable feedback from sponsors when they're back in their own country erm which had to do with the degree to which the course has made a difference. But that's that's usually informal. We do a course report which erm happens at the end of every course and in that course report we kind of evaluate what's gone on in the course and the course report always includes

a section which summarises the questionnaire which is administered at the end of the course. And erm that course report is often erm the starting-point for a discussion between us and the sponsors.

What kind of internal discussion takes place as far as the course report is concerned ?

27. Do you mean after it's been written [*Hm hm*] or er ? It very much depends on whether the course is part of a pattern that's meant to be repeated or whether it's a one-off. Erm if it's a one-off what we tend to look at in reports are those issues that might have staff development er consequences like erm if a particular tutor has obviously not sussed out what the course is about or what's required and then we pick that up and work with the tutor concerned, If it's a course that's likely to be repeated erm we're probably going to study it more carefully talk about it and erm and try to ensure that whatever mistakes we made on that course would not be repeated. And that of course also has a staff development dimsension in that erm the people concerned the tutors concerned are erm made conscious of things they have to work on in their own training style or whatever.

Do you think there are any erm evaluation procedures which have evolved over the time that you've been there which you're er which you're proud of which work extremely well ?

28. I think the six-month-on one works best of all...

With the contract groups the closed groups is there any opportunity to visit the students in the field to assess for yourselves the extent to which they've been able to well to what extent they've assimilated er what you've done with them and are able to apply it ?

29. Sometimes there is sometimes there isn't ....

Do you deliberately try to cost it in when you're asked to tender for a course ? Or is it not a primary concern ?

30. I think what we try to do is read the course from the point of view we read the potential. If we're aware that the course that we're tendering for has development potential and erm might lead somewhere then we would try to take account of it. And equally if we're aware that it is going to be the last course of its type erm we might be evaluating fairly rigorously internally but not give the same emphasis to external evaluation. [*A pragmatic perspective, really.*] A pretty pagmatic perspective.

What about pre-course evaluation, needs analysis and that kind of thing ?

31. Erm there's a big difference between open-enrolment courses and closed courses. The open-enrolment courses we get papers on all the participants, the application papers the degree certificates the c.v.s and so on and erm they provide a starting-point but we find very often we're more confused than we would be without them because everybody comes from such a different background erm so we use we may very well use that information from before the course to open up the debate about different needs on the course. The

closed group type of course it's extremely valuable to have pre-course documentation available to help us plan ....

What kind of things are you interested in getting information about ?

32. It depends on the aims of the group ... their experience, their language level [*Hm*] the expected professional roles of participants when they return ....

Going back to the Master's courses.... You haven't mentioned an External examiner.

33. We do have External Examiners .... They're extremely valuable to us.

If you could institute other forms of evaluation than those you're currently using what would they be ? I mean would you feel the need to introduce other procedures ? Do you feel that evaluation is important enough to merit [Yeah] *the kind of attention that I'm suggesting ?*

34. Yeah I think evaluation is extremely important. We think it's very very important and take a lot of note of what we get in the way of feedback from evaluation both formal and informal but I'm not sure that there are other things that we would necessarily want to do. Perhaps one of the things that we could benefit by would be to erm something you alluded to earlier on during a course for a colleague who's not teaching during that course to come in and could find out what's going on in the course and feed back to course tutors. I think that could be one thing that I I would possibly do that would be useful to the growth of a course while it's going on. All our efforts at evaluation have been definitely influenced by manpower problems and time problems [*Hm hm*]. We do as much as we can within the limits that we've got ... What I would find counterproductive would be if the data we were getting from the evaluation erm was more than the tutors' team could cope with and sometimes I think we get very near that.

INFORMANT 3

[Discussion of two Masters courses run by the institution] *At that level, do you have any difficulty with the term 'in-service' in deciding whether a course is an in-service course or something else ?*

1. Well, on the face of it, I wouldn't have thought there was any problem in deciding that because erm although again I'm not directly involved in those programmes erm but the minimum entry requirements are that people should have had at least three years of teaching experience erm I'm not sure that they ask for anything beyond that er well they must have a first degree or equivalent. Now if you see in-service as meaning having alreday taught erm yeah for erm well a a reasonable period of time then I would have thought there was no question that these are these are in-service.

Is that how you see it ?

2. Ah well, now this is a good question. You make me realise that I've not really thought about the meaning of the term. Erm no, I would in fact say that it really implies that people have had some kind of initial training as well as prior experience and and indeed that that well strictly speaking I suppose you could say that the experience is on-going and er and but yeah that would apply I think. It needn't be in the same job it can be between one job and another. Erm, so as as we talk it makes me think erm whether I do have an adequate concept of the word in-service, but yes I think it should include both at least both of those things: a certain number of years of having taught with an intention of going on to continue to do so but also already got some had some kind of meaningful initial training and this is in-service training is therefore something in addition to that something that builds on those initial experiences and qualifications. [Hm] So I think that the the MA programme is an in-service course in the sense of erm the teachers already having done some teaching and are likely to go on to do some more but not at all necessarily in terms of them having done any prior training in English language teaching. [*Hm nm*] Erm however the another one of the award-bearing programmes the the other one we have is at the undergraduate level and that's actually run jointly by the Department and ourselves and [LOCAL TEACHER TRAINING COLLEGE]. That's a B.Ed.TESL programme and that one erm I I think yeah I mean it is intended to be INSET and I think it is genuinely INSET in the sense that I've been using the term so far but because the people on it are Malaysian students who've all done a teacher training college course already and have been teaching for a number of years so erm that one is more genuinely INSET in that respect...

How far do the evaluation procedures that you use - evaluation of the course - erm vary from one type of course to another ?

3. ... the intention is erm ... to follow pretty much the same basic pattern. ... The basic idea of having erm erm you know formal evaluations at regular intervals and erm making use of that information during the course and at the end of the course and before the next similar course I mean that that's the sort of basic pattern that applies for all these courses regardless of whether they're certificated or not ... we use the

same basic approach to evaluating them.... summer courses ... tend to be closed-group ones, two-three weeks in duration and with them they all have an end-of-course evaluation but if it they were only two-week ones like the Greek one I did this year we didn't do any evaluation formal evaluation except right at the end but the group we have from France that I've also been involved with that's a three week course and we've usually done a brief evaluation at the end of the first week and then at the end of the course a more thorough one. But just given the shortness of them and also the fact that erm since they are courses which are which have been going for some time with a very similar clientele [*Yeah*] there's not the feeling of I think of needing to dig you know much deeper than that er and so on to in terms of our evaluation procedures...

Can I just pick up on two things that you mentioned in relation to the short courses ? [Hm] *One that because they've been running for some time with a similar clientele* [Hm] *you don't feel the need to revise the evaluation procedures very much* [Hm hm]. *And secondly you implied that they were short* [Hm] *which seemed to say you know either there isn't much time for evaluation* [Hm] *or it doesn't seem worth* [Hm] *evaluating very rigorously...*

4. Yeah, well, I think that I should add that as far as these courses go which er we've been getting for some time you know with similar personnel another factor is that the same staff have been involved with them [*Hm*] for some time as well and so erm in in I think that makes us, yeah, that definitely makes us feel that although there's always something to learn and there are always things you should be watching out for erm and you know there's always something you want to change at the end of the day or before the end of the day it's not quite the same as not at all the same as really as dealing with a course which is much more novel in all of in any of these respects. So I think that wisely or otherwise that is the the view that we've adopted abou those courses. I mean there is also the question of informal evaluation as well as formal evaluation erm. The informal evaluation is well goes on all the time I mean it's there from the very beginning. I mean the (laughs) the very first session especially on the French course my colleague erm who deals with that course and has done I think since it started about twelve years ago is able quite unerringly in the first session (laughs) to spot (laughs) certain tendencies among certain participants which which sure enough you know develop er just as she predicts erm during the ensuing three weeks. So you know I think that erm this this is evidence to some extent of the fact that that we are reasonably on top of the course and and how we feel it's going to go and how the people are going to be and to what extent things will suit them. Erm but also it's it illustrates that there is a great deal of informal evaluation [*Yeah*] that is all the time taking place. Erm but coming back to the the second point about the shortness of the courses, yes, I think we do have in our minds at any rate I do erm a feeeling that um in a short course although it's certainly possible to erm do a certain amount of evaluation you basically you've got to have done it beforehand really the that I suppose is when you can't evaluate certain things beforehand obviously but erm to the extent that you can that's when you really need to do it because in a short course it's very difficult to have a major change of direction I mean everything's happening so quickly. You know, by the time you've thought about (laughs) making a change it's too late (laughs) the course

has finished erm. I mean, that's not to say it might not be very useful for the next time round but I think it does we do we do operate on the assumption that or I do anyway that in in the shorter courses erm if it's less than if it's two weeks or less [*Hm*] I suppose two weeks is the shortest in fact [*Hm*] that we have in terms of sort of formal and thoroughgoing and major evaluation it's often unfeasible. That's not to say that there isn't room for plenty of other you know more informal and ongoing fine-tuning type of evaluation. I think that would be how I would see it.

Right, right. One of the the areas ... in which there seems to be relatively little activity across institutions is erm what I'd call post-course evaluation - so say three months on or even later [Hm. Yeah] *Erm could you say something* [Yes] *about what you do there ?*

5. Yes, well we do very little. Erm I think that with some courses fortunately there is a mechanism for erm things like visiting the places where the participants teach erm and seeing how they're getting on and and although on the face of it it's often erm the purpose of such visits is to to give them further advice or something like that it always turns out to be the case that that it's very very useful for erm further course evaluation. We have this with the French course for example every every two years the coordinator is able to go to France and visit some of the teachers and see how they're teaching, talk to them about how things are going [*How long after the course has finished ?*] Erm well it depends which year which group it was. It might with one group I suppose it would be about eighteen months after whereas with another one it would be six months after. It's every two years you see. Erm erm [*And has that six months eighteen months been chosen deliberately as an appropriate time at which to carry out that*] No no in fact it's not our initiative at all [*Hm hm*] it it's really the French government erm and the British Council who over the years I think have just developed this policy. It's partly to do with the planning meeting the following year and I think it was felt by someone, probably the British Council when it was involved, to be a good idea to combine this with a bit of evaluation or fieldwork or something like that [*Hm*]. Erm so I wouldn't I don't think that and it certainly wasn't a conscious policy on our part and I wouldn't pretend for a moment that we've got erm any kind of meaningful planned you know long term follow-up evaluation follow-up programme or scheme at all. I mean, there are other courses like the ones we're involved in for [PLACE] which has had a considerable because that's such a long programme that's been going on for about five or six groups now, there've been a lot of visits backwards and forwards and since since [X] has been here as Director he's encouraged a lot more of this kind of thing. And I suppose yeah I mean I hadn't really thought of using the word evaluation for it erm but a great deal of it is really that when you get down to it. He he's very keen on if we have a major involvement in fact now when we had a visitor earlier this week he was he was from [PLACE] he was saying that you know for anything meaningful to happen between us and them it would involve as a minimum at the beginning them coming and seeing us which was that was a condition that had already been satisfied by the presence of this person but it would equally involve us having to go over there [*Hm*]. And erm I think the pattern increasingly is that that would just be the beginning of a series of exchanges er or visits in both directions which is I think rather different

from the pattern that has tended to be in the past of it being much more one-way, but that as also associated much more with open-entry courses [*Hm*] which makes it less feasible to have that kind of of follow-up in that way...[refers to past attempt to follow up] but that was very much through erm questionnaire [*Hm*] to participants. And I think we kind of gave up on that because it because of pressure of work I think, just inertia, but also because it didn't seem to be yielding - very low return-rate [*Hm*] - it didn't seem to be yielding a great deal of information. Erm, we have a newsletter that we publish every year or thereabouts and every time in it I ask for information from erm past participants about how they're getting on and so on, the kind of information that could well be used for evaluation purposes. I've had absolutely nothing [*Hm*]. Erm well nothing that I've really thought of as being [*Yeah*] the kind of thing that I would like to have [*Yeah*]. Anyway, so I think because of a combination of the sort of factors I've been talking about we've not a great deal of meaningful activity going on in that respect with the possible exception of some of these single-country projects that have come up in more recent years. [*Do you see trend in that direction ?*] Yes, I think so. I think I've thought of it as being something wider than simply evaluation erm but ... as we talk it makes it clear to me that evaluation is a major aspect of that, indeed.

I have two more questions that I'd like to fit in ... The first has to do with information that is available on participants as input to evaluation [Hm]. *For example, when people come to you they'll normally fill in an application form* [Hm] *or they'll write a letter or both of these things* [Hm] ... *Do you ever go back to that information ... when the results are published ? Do you use it in that way ?*

6. Erm not to any significant extent, I would say. I mean, a lot of the time we really don't we we well I think we don't have much say in the people anyway [reference to closed groups] we accept whoever they give us and I'm not sure if we're given a great deal more than just the names and where they teach and a few things like that basic biographical stuff rather than anything further [*It's probably more relevant to the Certificate courses the open-entry ones*] Yeah, but even those we do get more information there because they've got to fill out a proper application form and erm or we get the British Council type of information erm that again I think that I really find it quite hard to to to make much out of those forms initially and you know I think that makes me feel that later on erm it it would be you know I find that I learn about the participants from the participants and I don't really feel the need I keep comparing that picture that is unfolding [*Hm*] erm from them directly erm with previous ideas and future possibilities rather than going back to the paper which I find you know is often not very helpful in in even forming initial impressions [*Hm*]. Erm I think the only exception to that would be - this hasn't happened for a long time - erm if there was a clear discrepancy between an English language test score on that sheet and a student's performance I think I'd go back to check on that sort of thing [*Yeah, yeah*] but otherwise very rarely [*Right*] we do that. That's not to say we shouldn't do it but you know maybe we should be doing so but in practice it's it's very rare [*Right*].

Would you say that overall you're fairly happy with the pattern of evaluation procedures that you've evolved over the years ?

7. I think I'm reasonably happy with them as they happen during the course. I mean, I think that the system of collecting information, I think the kind of information we ask for, the frequency with which we ask for it, the what we do with it - because it's all summarised and discussed with the participants and you know we make changes according to how they and us feel about it all erm - and then the final evaluation and then the other ones are all used to make the report on the course and that's fed into the next course or other similar courses. So I think that that procedure is is is all right really because it's it's I don't think we're overdoing it I don't think we're underdoing it we're getting reasonable information erm and so on, but I think that in terms of any real idea of the value of the courses to the participants in their own teaching situations I think about that aspect of it we are far more in the dark about and really lack adequate information. And erm it's it's vital to have it and so we go on operating on hunches and guesses and assumptions perhaps quite unwisely erm I'm sure very unwisely in many cases because of the lack of information of that kind. Erm and in fact I think I would go so far as to say that it's impossible to get it. I think that's the point. And that in a way makes one could make one question the value of courses of this kind that are so removed from the erm context and which it is so difficult to get meaningful information about. I mean, especially the open-entry courses it's less of a problem erm as I was saying with the newer project-related courses I think things are much better in that respect. I mean, we have we try and make sure that people get to know that country and those people and their teaching situations very thoroughly and they stick with that course and they develop that knowledge erm but with the open-entry courses it's really a problem. I mean, we do there are pa there are you know exceptions to that as far as the some of the participants are concerned that people from [OVERSEAS INSTITUTION] and I and colleagues have been working quite closely with them but they only form one or two out of a group of ten or twelve and the other eight or nine we may never have none of us may have been anywhere near the situation that they they operate in [*Hm*] and I I think that that's a serious flaw in our operation. Erm erm and and I I don't know a remedy to it. The only remedy I can think of is the one that would talk me out of a job (laughs) unfortunately (laughs). In other words you know a great deal more of the training should be based in [*Yeah*] in their own country in their own school and [*Yeah*] what have you.

INFORMANT 4

[Post-course evaluation your initiative or a requirement of sponsors ?]

1. Well probably the first one I was really involved in was the French Ministry programme [*Hm hm*] and I suppose that was because that was a kind of ongoing thing that was developing I felt it was important to find out what was what effect the course was having because I found that [in?] the end-of-course questionnaires people tend to be terribly enthusiastic [*Hm*] about the value of certain components and then I found it quite interesting to I found it quite interesting to follow up to find how much was actually implemented in and if things hadn't been implemented why. So the follow up allowed me to continue the course development and I suppose because I'd started off doing that with the French one that carried tended to carry over [*Hm*] into our pattern of course evaluation and course design, a little bit like with the Hungarians because we've been very fortunate there in that there's sort of been ongoing evaluation people having opportunities to go to the country [*Hm*] and follow up [*Hm hm*] and I think that the course design has improved [*Hm*] because the other thing too is that the French Ministry programme has been running for 15 years and I think that the the composition and the quality and the teachers have evolved over those 14 years [*Hm*]. It's quite interesting to sort of look back at where we were (laughs) 15 years ago. [*Right*] So it's really to develop a course and see what what how we can improve on it ...

Given that this particular course has been going on for 15 years erm have the evaluation procedures changed over time ? Clearly you have modified the course in various ways but has the evaluation also changed ?

2. I I don't think the evaluation has changed very much because the procedures were generally to have ongoing evaluation, [*Hm*] end-of-course course questionnaire, ahm then follow-up visits to France to see observe what's actually going on in the classroom [*Hm*] plus and that is actually our has been set up by the British Council and the French Ministry that kind of follow up [*Hm*] evaluation but the follow-up evaluation questionnaires we sent out that was very much our our initiative [*Hm*] to try and reach the people who we didn't have a chance to see in class.

So was the erm was the questionnaire part of the the follow-up questionnaire part of the package from the beginning ? Were the visits part of the package from the beginning ?

3. No no. The visits er the visits came out of I think requests from us from the Course Directors in the UK [*Hm*] that it was quite difficult to evaluate the impact of the project in the beginning particularly because we didn't know what was happening in the French classroom [*Hm hm*] and we didn't really know whether any of the participants were actually implementing any of the materials - because what we did was quite a bit of materials design and production here [*Hm*] the idea being that the teachers would go back and try it out in the classroom and so the follow-up questionnaire came for us because we wanted to see if they'd actually used any of it or any of the techniques and whether they felt they'd been they'd worked and why not that kind of feedback.

And how much feedback of that kind do you get or have you got in terms of percentage of returns ?

4. Percentage of returns ... a surprisingly good percentage ...

Good enough for you to think that it's worthwhile doing ?

5. Yes yes yes.

And when you say "surprisingly" is that because you don't really attach much expectation to people [Well I think] *returning questionnaires ?*

6. Yes I think the first time I sent out the qestionnaires, because I know the French are quite busy [*Hm*], that you've got to get the timing right [*Hm hm*] and I think we were we played around a little bit with with the timing and that's [*Hm*] also critical because if it's if it's too too soon after the course you get erm you might get returns but nobody's (laughs) tried anything out [Hm] and if you leave it till later on in the year when they're really busy marking exams then you don't get returns but we found Easter was quite a good time ... the course would be in July [*Hm*] so we're talking about what nine months ...

... so you would send questionnaires out in the second term of their year.

[Reason for thinking that enthusiasm might not carry over into practice ?]

7. Because I think the the limitation I mean part of the the feeling we got was that French tend French teachers tend to work very much in isolation and the benefits of our course was we had people working collaboratively on materials [*Hm*]. You'd get people into groups who had particular materials they books they were using so working together on that so that there was this kind of enthusiasm of having worked together [*Hm*] to get the stuff together but whether they would actually use them in the classroom given the limitations of time and the need to get through a a fixed syllabus ... the Colleges teachers now have a lot more freedom and I'd say that now a lot more is implemented than was in the earlier courses.

How standardised is the observation that you carry out of the teachers who've been on a course here ? So you want to see what effects the course has had, you go into a number of classrooms [Yeah], *what are you looking for ? Are you looking for the same thing in every classroom ?* [Ah, just] How do you evaluate ?

8. How do I evaluate ? Um just um I suppose it's interesting to go into the classroom and see whether anything any of the techniques that we suggested are employed. Sometimes you go in and the teacher's just performing in exactly the same way as they always have done [Hm hm] a lot of mother tongue a lot of things happening in French and then you go into another classroom and you'll see somebody actually developing some of the ideas that were developed in [PLACE] and you might go into another classroom and see somebody's who's moved on and actually produced something [*Hm*] in France as it were [Hm] So I think I'm interested to see whether the course here has had ***any*** impact ... We found the interesting thing is that although French teachers within their own school don't collaborate very well one of

the products of this erm this networking thing is that quite a few of them get together [Hm hm] once every two months [*Hm*] for a kind of reunion and [...] exchange material ...

How can you be sure that what you see is different from what you would have seen if you'd been into that classroom before [We] *the course ?*

9. We can't ...

You don't get a chance to do a before and after ?

10. We don't ... only on one or two occasions. It's not the policy from the French end. Sometimes we get a chance to see teachers who we taught like three or four years [*Hm*] before. That's that's quite interesting. And I think on the odd occasion we've been able to observe one or two participants before but that's not normally the case.... I wish we could but it's really the financial restrictions... [Explains that the cost of the visits to France is borne by the institution and that, as far as she knows, the other centres operating this programme do not carry out visits of this kind, although all the centres are obliged to attend an annual meeting in France.]

How do you know that what you see when you go into the classroom if it does appear to draw on what's been done on the course isn't just being put on for your benefit as it were [Yeah] *as opposed to being you know something that's been fully integrated into* [Yeah] *this teacher's* [I know] *repertoire?*

11. I think often the reactions from the pupils tell you (laughs) tell you quite a lot as to whether they're used to this or whether it's something very novel. Ah again discussion with the with the teacher erm [Hm] usually. They tend to be because we it's quite a close group usually after the three weeks everybody's bonded quite well [*Hm*] and there is quite a degree of frankness [*Hm*] and er talking through the problems of the classroom situation [*Hm*]. Because obviously schools differ [*Hm*] and in some schools it's easier to implement new things [*Hm hm*] than others.

And how would what you gleaned from the visit of this kind feed into what you do here ? So you've discovered that some people do implement some of the ideas, some appear to, and some have even gone beyond them [Hm hm]. *So what ?*

12. We probably alter the course programme. I mean we feel or rather I generally feel that er it's this idea of working together and trying to build up networks.... In the old days we were very much more inputting [*Hm*] you know input slot, little materials slot [*Hm*] and then switch skill, but I suppose what happened as a result of of going into the schools and taking on board really the the facilitating role of getting them together producing things erm means that some people would be producing new stuff that'll get circulated and again when the st when the new people come in they can see what was produced last last time antd if there's anything they want to photocopy they can photocopy that and take it away ...

If on one of your tours of the colleges you erm you came across let's say three teachers who didn't appear to have been affected by the course here at all [Hm] *would you try to explore the reasons why that might be ?*

13. Yes. And I think [*And have you in the past ?*] I have in the past mainly through well talking first of all with the teacher and probably reflecting on how that person performed on the course [*Hm*] how involved involved they were and whether you could have predicted the ones that were going to erm implement or not right from the beginning ... as years have evolved I think selection processes improved at the French endand the teachers now have to pay their own travel ... in the early days language was a problem for a lot of the teachers and I think generally the linguistic ability of the teachers who're coming here has improved [*Hm*] again I think it was an age thing the younger generation seem to be much more on the ball [*Hm*] in terms of methodology quite a few of them have been on other methodology courses in this country [*Hm*] so obviously what we're now observing is not just a result of our course but a result of I suppose the French Ministry's commitment to to training [*Hm*] teachers of English [*Hm hm*] so it's probably getting a bit more difficult now to evaluate how much is as a result of [*Yeah directly*] directly our activity and how much is a result of workshops held in France [*Hm*] and other courses.

14. [There have been modifications over time, especially in relation to the language improvement component, which was in the main brief initially.] Within the earlier groups, the ones who probably didn't implement were the slightly older ones who were less confident about the language and less confident about the technology, about using a tape recorder, that kind of thing....

15. [Other courses] We do follow up but we don't have the opportunity to go and observe ... so the only follow up we do there is questionnaires and again the return is quite is quite good. They do tend to keep in touch. And often you get unsolicited feedback, which is nice. [Mentions information from teacher based in overseas country that former participants meet once a month.] ... with the open teachers' courses it's more difficult to monitor what goes on two or three years down the line.

16. [questionnaire responses again] I suppose we make quite a big thing about the importance of feedback to us [*Hm*] and that getting questionnaires saying how wonderful we are isn't terribly helpful (laughs) to [Hm] the course development ... and therefore I think we're not just going through the motions as it were ... part of the key is trying to impress upon them while they're here how useful evaluation is [Hm] by demonstrating how we implement what comes out of our evaluations ...

Which of the various evaluation procedures that you use ... would you say gives you most useful information ?

17. I find that diaries are really useful if you can persuade the students of their value and often I find that a two-week course is not long enough for some individuals to to grasp how useful that can be for us [*Hm*] because it gives you perpsective on well the students' sort of uninh sort of impression of what's gone on [*Hm*] that day ... [Refers to occasion recently

when she was preparing a tender for a group of a type they had last had two years before] I found that going back though the diaries actually helped [*Hm*] to bring back memories of what had gone on [*Hm*] and in a way were more useful than the post-course evaluations that we did ... because you could see how things progressed through [Hm hm] through the course and also if you've got twelve diaries you get different perspectives on what's actually going on. It would be really quite nice to get teachers to to keep diaries of what what they do when they go back and follow up ... [occasional students have not wanted to keep diaries because] they just didn't like reflection I think. [*Right*] And I think this whole business of reflective practice is something which is fairly alien still to a lot of teachers and they kind of see it as intrusion some kind of intrusion [*Hm*] ... I've tried different approaches and sometimes feeding in an article on the value of diaries [*Hm*] - this went down like a dead duck (laughs) last time I used it - erm showing samples of different diaries erm talking through them how you know what they would perceive as diaries. Tried doing sort of group group diaries [Hm] erm but then and again that's sometimes quite interesting to to see how people influence each other in terms of how what they put down [*Hm*] and how they perceive it ... But I must say it's been much more on an *ad hoc* basis rather than a systematic approach.

Do the participants know that you're going to take them in at some point and will be looking at them for purposes of evaluation ?

18. Yes.

They do? You announce that at the beginning ?

19. Yes. [continues]

Would you say that you're satisfied in general with the systems of evaluation that are set up ?

20. Ahm I think we need to reflect a lot harder on what we're doing and try and make it more more systematic because the danger of what often happens here with a course - not like the Hungarians and the French but the one-off courses [*Hm*] - I think possibly procedures are just used because that's what they did that's what was done last time rather than thinking [*Hm*] well perhaps we could just concentrate on one one form of evaluation like give diaries a go or [*Hm*] because sometimes I think the students end up feeling can end up feeling overevaluated overevaluated (laughs) [*Hm*] and particularly the Hungarians often get really annoyed because they kind of feel that this evaluation-reflection thing is a sort of indication that the teacher doesn't quite know (laughs) where the course is going and er so it's something I think probably we do need to sit down and [*Hm*] have a better think about. Because the mixed the mixed groups every group's always different and er I suppose when we've got monolingual groups they force you to think a bit harder about what you're doing [*Hm*] because well they come from the same backgrounds ... that's probably why we concentrate more on it than we do with the open teachers' courses ... I think we give more attention to evaluation ... It's probably the time thing too ... the shorter courses are only two weeks [*Hm*] and I think that you're trying to probably pack so much into the into the two weeks you think that too much

time given to evaluation is perhaps perceived by the teachers as a bit of a waste of time. [*You mean the participants ?*] The participants, sorry, yes.

INFORMANT 5

The Dip TEFLA ... is assessed externally [Yes] *and validated externally* [Hm hm]. *Erm how interested are you as far as your evaluation of the course is concerned by the results ? Do you to some extent say 'Well, we must be doing all right because the results are good in our terms'? Do you lean on that as confirmation that you're doing a good job ?*

1. I think for the most part we do.....

Would you say that because you have this kind of external measure of the success of the course that you feel under less pressure to evaluate the course for your own purposes using your own means ?

2. Yes, I think that's fair. There isn't time [*Hm*] to evaluate as we go along [*Hm*]. Erm I know that sounds (laughs) unscientific but it's perfectly true [*Hm*] .... The evaluation that we do ... is more to do with the feedback [*Hm*] working out finding out from the s the s participants how they're erm feeling at any particular stage and whether they need more or less [*Hm*] depending on just where they are [*Hm*]. So we don't evaluate formally as we go along ...

And do you carry out any kind of post-course follow-up to try and trace people afterwards ?

3. No. I think it's important to mention that on this course the trainees have an enormous amount of individual attention from tutors as a result of returning essays, feedback on practicals and workshops plus individual personal tutorials, about one hour each three times during the course. No formal written evaluation, but a system that encourages continuous feedback.

What criteria are used to evaluate open-entry short courses ? What are you measuring ?

4. Satisfaction.... Our open-enrolment groups erm for the most part are coming as individuals [*Hm*] self-financed erm during a holiday period [Yeah] and for the most part what they want is an update on techniques and methods or if it's language development they just want to improve their language [*Hm*] find out what's happening in England today. So no there is no formal evaluation beyond the student the client satisfaction.

Are you happy with that ?

5. In a two-week course, yes. It'd we'd find it very difficult to do anything else. Because the first week a lot of it's just trying to settle them in, find out what they want, see that help them find out what they want in relation to what we can provide [*Hm*]. Erm and that does take a good week to do that [*Hm*]. And the second week then is providing the best one can and to give them opportunities to do more things if they want to.

[Closed-group courses] *Does the sponsor ... determine how you evaluate ?*

6. No, I can't think of any example where they said 'This is what we want you to to do [*Hm*]... Erm no [specific closed-group course mentioned] ... we are required to test them with ARELS [a laboratory-based test of listening and speaking] ... and they have to take it again before they leave...

What other measures do you use on that course which in a sense you've taken decisions on yourself ? I mean, that that criterion would be improvement in language performance obviously.

7. Yes, I'm just I don't think that there are any more erm. For evaluation?

Hm. Programme evaluation. Course evaluation.

You were the [JOB TITLE], weren't you ? [Yes]. *And one of the things that you obviously did was to debrief Course Directors* [Yes] *if you hadn't been directly involved in a course. So the substance of your conversation presumably would have been 'Well, how did it go and how could it be better next time ?' and so on* [Yes] *and so essentially you were evaluating the course*

8. I'm really hesitating over how to answer this. In that job I saw CDs informally twice a day. This is when a lot of information changes hands and advice is given on whether or not to spend more or less time on a topic or to change the type of input depending on group dynamics. I don't see this as evaluation but tutor support. However it has implications for evaluation. The reality in relation to the course I was talking about before is that trainees have an extremely good course. Over the years course content has consistently scored 'Excellent'. When we discuss how to improve it we find the discussion involves personalities and how they react to types of input or material. This you can't prepare for but it is only by having group harmony and a sense of working together that the course can really succeed. Hence the daily meetings of the Coordinator and the CD in which the trainees' reactions are informally evaluated by the simple yardstick of 'Did it work or not ?'. Their views are constantly invited. Tutors and trainees have a very informal friendly relationship - they usually have coffee and lunch together.

Erm what did you draw on as as evaluation evidence for that course, for example ?

9. End-of-course questionnaires. [*Hm hm*] Erm and just the knowledge of the course as well.

So again it was participant satisfaction [Yes] *that you were trying to get at through the questionnaires.*

10. Yes yes.

Anything else ?

11. No. That's a four-week course. The sponsor comes once sometimes it's twice but on the whole it's once during that course ... and he or she will see every participant and then feed that back to us [*Hm*] so there's another

evaluation [*Right*] going on. [The institution does not monitor the effect of the course when participants return home.]

[short courses] *You probably feel disinclined to allocate time within the course class time* [Hm] *to any kind of evaluation* [Hm] *other than for filling in questionnaires or something like that* [Hm].

12. There's a weekly evaluation well class discussion [*Hm hm*] usually on a Friday afternoon when they discuss what they've done during the week.

So it starts as a review [as a review to feed into the following week's work.] *Right. Erm is it just descriptive or is it evaluative and then ... ?*

13. It's descriptive.

It helps them to sort of place things.

14. Yes, but also yes it helps them but it's also helpful for for the teachers to know where to go the following week [*Hm*]. That's [*But you're not actually soliciting comment on how people felt about what happened or*] Yes. [*You are ?*]

15. Yes. Oh yes. Yes. So if they actually say we don't like that particular kind of lesson [*Hm*] or don't do that again then that lesson the following week will change...

[*Participants asked to prepare for evaluative discussion ?*]

16. There are different ways of doing it.... Quite often it's done in groups small groups [*Hm*] feeding into bigger groups.

Hm hm. Would there be prompt questions of some kind ?

17. Yes. Yes...it could be open, the teacher getting ideas from the class.... It does it does work erm and students do come up and give say what they want and what they don't want particularly if they're only in [INSTITUTION] for a short time ...

You don't just get responses from the vocal minority [No] *in that sort of open* [No no] *setting ?*

18. Well in my er having just done it just this week I said 'No' but I can't swear that would be the case [*Hm hm*] because I'm sure that you're right that when there's a biggish group then there will be people who don't know what well don't know what to say [*Hm*] or feel too embarrassed [Hm]. That's not a good situation [*Hm*]. Erm what er in a lot of the teacher training groups is setting up small groups and then whether they're set up so that you've got the vocal ones together or the quiet ones together it's again it depends on the Course Director [*Hm*] but that's one of the methods the techniques rather that we use [*Hm*]. Erm sometimes putting it on a poster so that everybody can see or just having one person feed back to the group and the whole lot going on the board erm or using little labels and putting things on labels of things you want to change and sticking this on a poster ... Those are the ways that I know are used.

Does it ever happen that the tutors who're running a course or directly involved in it ask somebody else to come in and [Yes] *hold that discussion?*

19. Yeah, yes it does.

Is there anything that motivates that decision rather than one of the other options ?

20. I think it's personality... If they think that there are trainees in the group who might be difficult or who might not speak say what they think to <u>them</u> then they'll ask somebody else to come in.

Hm. And that happens quite often.

21. Yes. Yes.

But it's not a policy.

22. ... I think it's a norm now [*Hm*] whereas but I think this is fairly recent.... It's the sort of thing that began slowly and we now do that with all our courses. We have the senior tutor come in for the evaluation session at the end.

Hm. Is there a reason for that ?

23. I think that it ... it's just easier ... it's to do with feelings and not hurting feelings.

If you were in that position of going in to a colleague's course [Hm] *and taking the comments would you would you consult the colleague first about the sorts of direction that they might want the discussion* [Yes] *to ... I mean, the sorts of information that* [needs to be elicited, yes] *that one might be interested in getting ?*

24. Yes for the sort of grey areas where you just don't know what's likely to to how they felt [*Hm*] to channel that erm but also to to know who are the loudmouths within the group erm and to try and make sure that they that they don't dominate to the point that the other ones don't say what they want.... There has to be about half an hour's discussion beforehand...

What does one do either individually or institutionally with the the findings ?

25. Oh I don't know (laughs) to be honest. For a weekly ... evaluation then obviously it's fed back to the Course Director who then can do something about it for the next part of the course. [*Can do.*] Can do. Or may not want to depending on just what it is that's coming out [*Hm hm*]... Sometimes we actually move people as a result of the Friday afternoon talk [*Hm*]. Erm it can have that sort of repercussion [*Hm*] and affect the group in that way but otherwise it may just be talking to them that these are our constraints this is what we have to work within [*Hm*] and trying to get the group to come together but on the whole we don't have to do that very much....

What would you do with end-of-course evaluations ?

26. Well the end-of-course evaluation has to go to the central organisation ... I don't know what happens to it. Put in a filing cabinet and lost for ever. .... [Then] there's the standard questionnaire that we do on the teacher training courses which we keep [*Hm*]. That's much more to do with the sessions [*Hm*] which sessions they preferred which ones they didn't erm and then of course there's the discussion as well that's just notes that doesn't that will go into the course file for possible help in preparing future courses.

So the Course Director doesn't write a report ?

27. Yes, and the Course Director writes a report.

And who sees that report ?

28. Erm Director of Studies Head of Studies and senior tutor....

What happens then ?

29. It's filed. And then when the next year this course is being prepared the next Course Director who may not be the same person as the previous year refers to it and (laughs) if there is wisdom to be gleaned from it they then (laughs) try and do something.

So there's no discussion between the people who receive the report and the person who wrote it.

30. Yes, in the debriefing. Right. Okay. While a course is ongoing there is the evaluation with the group. The course finishes [*Hm*], the Course Director writes the report [*Hm*], which goes to the Senior Tutor and to Head of Studies and there's a debriefing involving the Senior Tutor and CD, but in the summer the debriefing tends to be it's October [*Hm*] before [*Hm*] and quite often the debriefing is not er it's with all of the people who did those summer courses.

It's a general meeting.

31. The teachers' courses.... I say quite often. It's time and who is going where and when. Tutors may complete a course on a Friday and then go on leave for four weeks ... By the end of the summer a CD may have taught on four courses and due to lack of time not have been debriefed. [*But decisions are not taken at that time about changes* [No] *that might need to be made in the future. They're left until*] Oh yes, no that's not true. If there's anything that's badly wrong then yes a decision will be taken at that time because it will Head of Studies will if there seems to be a pattern that things aren't working the D there will be another meeting with the Head of Studies and the teachers concerned with those courses and and that will then lead into erm meetings with Marketing to revamp [*Hm*] the kind of course that will be needed for the following year. During the rest of the year the debriefing will usually follow the week after the course has left. The CD may by that stage be a tutor - not a CD - on another course and be less pressured. However in cases where a course report has to be sent to a Ministry or sponsor within six weeks of the course ending the debriefing has to be fitted in as soon as possible.

[closed-group courses] *On some occasions you have had a chance to follow up ... whose initiative did that derive from, yours or that of the sponsor ?*

32. Usually the sponsor. Often a combination of both. Erm if ... it depends who's got the money to pay. For the most part we don't have the money for people even in the past we didn't have the money for people to go and do evaluation of little courses ...

Would you have tried to push [Yes] *a sponsor to incorporate that ? ... And how valuable did you feel it ?*

33. Extremely valuable. Erm the biggest the longest-running project that we've still got is the erm [NATIONALITY] teachers.... There was a point towards the late eighties when we suggested that it'd be quite nice if some teachers could go some of our staff could go out [*Hm*] and just look in the schools and just see what was happening [*Hm hm*] so erm two members of our staff over a three-year period went out to see classes there and to give er talks but they paid...

How did that erm feed into to your evaluation of the programme ? What did you learn from that little exercise ?

34. I think the main thing that we learned from it was just sort of what the level of the trainees really was. I mean what the teaching situation really was, what was actually possible, what was not erm because if you're working in an English environment or British environment you can get carried away with [*Hm*] what is possible [*Hm hm*] and it's not until you actually get into the schools and just see a forty-minute period [*Yeah*] what you can achieve in a forty-minute period er and what the the erm I say the trainees because we didn't weren't seeing the trainees because I went on one occasion erm we just saw teachers [*Ah, you didn't see your former*] No no no I saw one of them [*Ah ha*] I only saw one.

Because that would have been the real test [Yeah yeah]. *What were you looking for when you looked at this individual ? What would would you have been looking for if you'd been only focussing on former students ?*

35. I'd have been looking for how much they'd erm learned from the experience of being in England, whether it was actually possible to do what we were suggesting in their schools even.

So would you have been looking at methodological practices as it were, the ways in which certain things were done ?

36. Hm and how it could fit in with the coursebook.

How would you know that what they'd be doing when you observed them was different from [Dunno dunno] *what they'd been doing* [*haven't a clue*] *before?*

37. I just wouldn't know wouldn't know at all [*Hm*] would just have to go by their word [*Hm*]. I mean you can tell up to a point as to whether or not the students are responding in a surprised way (laughs) or [*Hm*] in a way that

they just don't seem to know what to do [*Hm*] and you can tell from that [*Hm*] erm but otherwise you can't really. Actually, we do know that when this group first arrived the Ministry was about to introduce a new "communicative" coursebook into the secondary schools. We were helping with the transition and the move away from a traditional text-based coursebook that used translation.... When I observed the classes I was able to see how the methodology was being applied and to talk to the teachers about the new materials.

It seems to me you know that the real question about in-service provision and its effectiveness is you know what do people do [when they go back] *subsequently* [I know] *not necessarily immediately but* [No, I know] *over a period of time and* [Well] *at what time do you*

38. I went to [COUNTRY] for a week-end seminar (both laugh) with erm forty of the trainees who'd come to us they'd been coming to us for over a period of at that stage about five years for four-week courses on language development and er methodology it was a really basic methodology [*Hm*] because they're restricted in what they can do by the coursebook in [COUNTRY] erm and went back and had a seminar with what had they actually been doing since they went to England and nobody had been doing anything except for one. Only one out of the forty had actually tried to introduce anything new into the classroom. At least they were honest about it [*Hm*]. They said erm hey just said it wasn't possible.

So how did you feel about that ?

39. It was what we had suspected quite honestly. Erm the one who had been using English all the time the other said no they couldn't use English all the time in [COUNTRY] this is secondary school teachers at least there was one doing it out of forty (laughs) that's a start [*Hm*] and erm

Was he doing it as a result of the course ? [Yes] *Ah ha.*

40. And he was very enthusiastic about English and erm he just loved it [*Hm*]. But the others didn't (laughs) and so they were taking the easy option [*Hm*] and he had a vile teaching programme anyway I mean they had a very very heavy teaching programme diabolical books so it was hardly surprising (laughs)

Did it force you to think again about what you did ?

41. Well, what we did we actually erm changed the focus of the the course erm to be predominantly language development [*Hm*] erm it was originally a 50-50 split and we we shifted it to 60 language development 40 methodology because we were required by the Ministry to do methodology, which they couldn't use [*Hm*] but within the methodology we made it predominantly language with just some

So what was your rationale for that ?

42. Because that was the most useful thing for them to improve their own language because they were they were really low [*Hm*] they were low intermediate low to middling intermediate. At least if they were better

[*Yeah*] at least when they were using their coursebook they it a slight improvement on what they were doing before (laughs).

Did this particular experience cause you to think more generally about other courses that you were running, closed-group courses and whether in a sense a focus on language may be more easily transferable [Yes] *than methodology ?*

43. Up to a point. With often it's it's the Ministry or sponsor who will say 'This is what we want you to do' [*Hm*] and your hands are tied up to a point then [*Hm*] but most of our other groups are mostly from the West ... and we've got people going out to schools there [*Hm*] ... and so we've been getting groups from there and so we we have the knowledge of what the schools are like what their coursebooks are like what they're allowed to do [*Hm*] and what they're not allowed to do and that makes a very very big difference.

So in the case of the [NATIONALITY] you were relying on information from the Ministry which [was very limited]. *I was going to say was it it wasn't misleading it was just limited ?*

44. No no it was limited. We got a erm a kind of a leaflet handout from the British Council originally but it didn't really say it didn't really give us the information that we needed erm and when we were trying to er select the trainees were trying to influence the selection of trainees we found that the British Council were unable to get the Ministry to budge in any way. What was happening was that we were getting trainees who were friends of erm other people in the Ministry [*Hm hm*] and they weren't one on one occasion wasn't even an English teacher [*Hm*] and this made us very very angry [*Hm*] ...

Do you feel with hindsight that you asked the right questions ?

45. Yes, but how they were translated is something we don't know because we know all the questions that we had were then translated.... But we had a very big [NATIONALITY] project and that was totally totally different.... and there that's how a project should be with erm trainees coming to us first for language then they went back to their country for initial training in methodology then back to us for a three-month brush-up [*Hm*] on English erm you know this is how they do it in England to get enthused about it, back to the schools where they were monitored in the schools and we went out and actually worked with them in their schools for four months sorry four weeks at a time [*Hm*] and saw them teaching and had workshops with them in that every week and it was totally different [Hm] it was a totally different experience [*Hm hm*] and this went on for several years .... They're coming back to see us even now, because they're coming some are coming to do MA courses [*Hm hm*] so there is er informal follow-up going on but not as such evaluation at this stage. But then going back seeing what was happening in the schools seeing how awful conditions were [*Yeah yeah*] ... Post-course evaluation wasn't just post-course evaluation erm it was preparing for the next stage of another course [*Hm*] so yes that was real evaluation because you were seeing the results of one lot of input and how it fed into another rather than something that was static.

Were you seeing the people who would then come to you next ? Were you seeing the new [Yes yes] *cohort as well ?*

46. But it didn't really happen on any other course. With us a course leads on a Friday afternoon and that's the end that's [*Hm*] that's the last we see of them. No it's not very satisfactory we haven't a clue whether erm what they have learnt is is going to be used if it is at all useful at all erm whether it was even interesting for them [*Hm*]

[Problem of follow-up: difficult to make contact with trainees after they leave] *Have you thought of trying to follow up or tried to follow up and given it up or has it just never been on the agenda ?*

47. No, it's never been on the agenda. We've never erm to be honest we've never had the time we're always looking forward I mean this is the nature of any organisation I suppose [*Hm hm*] er and again what's its value what's it going to do what's its purpose ? In fact we've just finished and it's finished. If we don't get another group like that there's no point in following it up

What about the open courses ? Those are things that you would expect to run on an annual basis ?

48. Yes yes a very good point. But we've never done it. Don't think there's energy erm and always thinking ahead rather than backwards ... As you say we run these courses we should be following them up [*I didn't say that (laughs) you said that*] No, I said that. Well that's my yeah that's how I feel [*Hm*] in an ideal world but it's not possible

49. [comparing evaluation instruments] The end-of-course questionnaire tends to be the most useful ... Some people just fill it in very rapidly and they don't think very hard about it others do take it very seriously and they if they have any real concerns they usually put something in [*Hm hm*] at some stage in their questionnaire ... in terms of actual detailed information we'll get more there than in it's difficult to say it depends on the group and the group dynamic. With the discussion [*Hm*] the plenary or little group discussions with some groups you get a lot back and with others very little

[feedback from sponsors] *Do you find that helpful ?*

50. It depends what they want to tell us [*Hm*]. You just don't know how much they're saying t just to be polite or whether they are genuinely happy erm ... the feeling has been yes they are telling us wh how things are because that's how we think things are but if there were a problem that they'd been told we can't be sure that we'd be getting that ... so there's always a slight reservation are they are we being told everything ...

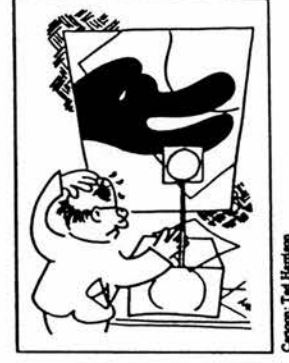
Summing up... what would you say the purposes of evaluation are, within the [INSTITUTION] ?

51. Information for the next lot of courses (laughs). Usually it's not possible if there's a short course it's often very difficult to make changes [*Hm*] er particularly if as can sometimes happen they're not happy

with the teachers ... fine-tuning for short courses for the following week, for courses of more than two weeks say up to four five weeks informative as to how the group can run best in the weeks that are left [*Hm*] erm getting them to work more as a group to see things from the organisation point of view not to sort of dream of something that is just not at all possible [*Hm*] so being realistic about what they've achieved in the time and I think finally there's in the long run is what can be done for future courses er if big changes have to be made ...

APPENDIX 5.1

Pre-seminar questionnaire


Cartoon: Ted Harrison

From: Ian McGrath
To: all ML tutors

USING THE OVERHEAD PROJECTOR

****

I have been asked to lead a workshop for ML teachers on the use of the OHP. This will take place:

Friday, February 19 (13.30-15.00).

I should be grateful if you could take the time to answer this short questionnaire and return it to my basket by:

Friday, January 29.

****

1a. Have you ever used an OHP in your teaching ?
Tick the box [ ] that applies.

a. No, never. [ ]
b. Yes, once or twice. [ ]
c. Yes, occasionally. [ ]
d. Yes, frequently. [ ]

If you answered (b), (c) or (d), please go on to Question 2.
If you answered (a), go on to Question 1b.

1b. So you have never used an OHP. Why not ?

a. I've never felt the need. [ ]
b. I don't know how to use an OHP. [ ]
c. There isn't usually a machine available. [ ]
d. I don't know where the transparencies/pens are kept. [ ]
e. I don't like using machines. [ ]
f. (other) ...

If you answered (b), (c), (d), (e) or (f), go on to Question 4.
If you answered (a), go on to Question 1c.

1c. So you don't feel the need for an OHP. What do you use instead?
Tick all the boxes that apply.

a. the whiteboard [ ]
b. prepared pictures [ ]
c. photocopies [ ]

Now go on to Question 4.

p. t. o.

2. How many times have you used the OHP so far this term ?

a. not at all [ ]
b. once [ ]
c. more than once [ ]

3. For which of the following purposes do you (normally) use the OHP ? Tick all the boxes that apply.

a. writing new words and phrases during the lesson [ ]
b. writing students' errors during the lesson [ ]
c. exercises prepared before the lesson [ ]
d. answers prepared before the lesson [ ]
e. answers given during the lesson [ ]
f. reading texts photocopied from newspapers, etc [ ]
g. gapped texts (e.g. songs) [ ]
h. pictures: e.g. photocopied cartoons, picture stories, etc [ ]
i. maps [ ]
j. pictures drawn during the lesson [ ]
k. questions/instructions for an activity [ ]
l. games [ ]
m. crossword puzzles [ ]
n. (other)

4. Tick the box(es) that describe your attitude to the OHP.

a. Don't really see how it can help. [ ]
b. Haven't got time to prepare materials. [ ]
c. Might use it if I knew how to. [ ]
d. Have had bad experiences with OHPs in the past. [ ]
e. Useful substitute for board. [ ]
f. Useful supplement to board. [ ]

If you answered (a), (b), (c) or (d), come to the workshop for information, advice or just to listen to what other people have to say.

If you answered (e) or (f), you're probably a frequent OHP user. Why not come to the workshop and share your ideas with other tutors ? You never know, you might also pick up some new ideas yourself.

NAME ....

Thank you.

APPENDIX 5.2

Post-seminar questionnaire

POST-SEMINAR INTERVIEW concerning OHP use

NAME ....

1. How many classes have you taught, roughly, in the two weeks since the seminar ?

   (a) two - one per week [ ]
   (b) three to six [ ]
   (c) more than six [ ]

2. Have you used the OHP since the seminar ?

   (a) no [ ] - go to Q.6
   (b) once [ ] - go to Q.3
   (c) more than once [ ] - go to Q.3

3. What for ?

	Q.3	Q.5	Q.6
a. writing new words and phrases during the lesson	[ ]	[ ]	[ ]
b. writing students' errors during the lesson	[ ]	[ ]	[ ]
c. exercises prepared before the lesson	[ ]	[ ]	[ ]
d. answers prepared before the lesson	[ ]	[ ]	[ ]
e. answers given during the lesson	[ ]	[ ]	[ ]
f. reading texts photocopied from newspapers, etc	[ ]	[ ]	[ ]
g. gapped texts (e.g. songs)	[ ]	[ ]	[ ]
h. pictures: e.g. photocopied cartoons, pic. stories	[ ]	[ ]	[ ]
i. maps	[ ]	[ ]	[ ]
j. pictures drawn during the lesson	[ ]	[ ]	[ ]
k. questions/instructions for an activity	[ ]	[ ]	[ ]
l. games	[ ]	[ ]	[ ]
m. crossword puzzles	[ ]	[ ]	[ ]
n. (other)			

4. Do you plan to use it again in the near future ?

   (a) yes [ ]
   (b) no [ ] - go to Q.7
   (c) not sure [ ] - go to Q.

5. What for ?

6. If you haven't used it, do you plan to ?

   (a) yes [ ] - go to Q.3
   (b) no [ ] - go to Q.7
   (c) not sure [ ] - go to Q.7

7. Can you say why you haven't used the OHP yet ?

(a) didn't feel the need [ ]
(b) don't feel confident [ ]
(c) no machine available [ ]
(d) didn't know where transparencies/pens were [ ]
(e) don't like using machines [ ]
(f) ....

....

8. Has your attitude to the OHP changed as a result of the seminar ?

(a) yes [ ]
(b) no [ ]

If YES, how ?

(a) gave me confidence to use it [ ]
(b) now see it as useful substitute for board [ ]
(c) now see it as useful supplement to board [ ]
(d) ....

....

....

9. Have you dipped into any of the readings suggested at the seminar ?

(a) yes [ ]
(b) no [ ]

If NO, why not ?

(a) didn't feel the need [ ]
(b) no time [ ]
(c) couldn't find them [ ]
(d) other

10. Did you speak to any other ML staff about the seminar ?

(a) yes [ ]
(b) no [ ]

If YES, what did you tell them and how did they react ?

11. If the seminar were to be repeated for staff who missed it, can you suggest any ways in which it might be improved ?

APPENDIX 6.1

Original evaluation framework for TLE

The instruments and procedures listed below have been abstracted from McGrath (1982) and categorised along a time continuum; where appropriate, a rationale has been included.

EVALUATION PROCEDURES: TLE 1982

Pre-course evaluation

1. ***pre-course questionnaire*** to elicit participants' professional background, potential familiarity with recent EFL theory, current practices, perceived needs and wishes (input to course design and baseline indicator of awareness and attitudes)

2. ***framework for participant formulation of individual language learning objectives*** completed on first morning of course (encouragement to self-direction and basis for subsequent consideration of extent to which these had been achieved)

In-course monitoring

3. ***individual tutorials*** during timetabled self-study sessions (designed in the first instance to encourage introspective reflection on course activities, the course process and as 'a point of formalised personal contact between participant and tutor: an opportunity for the latter to answer questions, deal with problems, and make the participant feel that he or she counts as an individual' (McGrath, *op.cit.*: 20))

4. ***rapid assessment forms*** (RAFs), completed daily and handed in weekly. Participants were asked to rate each session for value and interest separately using a scale of 1 (= useless/boring) to 5 (= very valuable/very interesting) and were asked to comment on the same sheet on any rating below 3 (= satisfactory). ('Detailed feedback from participants will be needed if any conclusions are to be drawn as to the success of the course, its component parts and the sessions within each component. And if feedback on individual sessions is to be sufficiently detailed (bearing in mind the blurring effect of time on memory) it needs to come as soon as possible after the session to which it refers' (McGrath, *op.cit.*: 34).

5. ***trouble-shooting sessions in Weeks 2 and 3*** ('the Forum session is envisaged as one in which prepared questions [suggested by participants] form a springboard for general discussion; the Integrated Activity will use small group discussion as a threshing-ground for individual problems' (McGrath, *op.cit.*: 29); see also Early and Bolitho, 1981)

6. ***weekly tutor meetings*** to review the past week (on the basis of the collated RAF results and individual perceptions) and fine-tune the next week: 'The sheets will be ... considered during a weekly tutors' meeting. This procedure has the obvious advantage over a single

final evaluation that it allows for a shift of emphasis or change of tack in mid-course, and for an exchange of information on participants' (McGrath, *op.cit.*: 34).

7. ***individual tutorials*** in final week

end-of-course evaluation

8. ***end-of-course questionnaire***, handed out on the penultimate day of the course. The first section of this, to be completed in readiness for the final day, is broadly concerned with level of participant satisfaction, but also seeks to identify sessions/topics generally perceived to be particularly useful or of little interest/value (see Appendix 6.5 for 1992 version). The second (more general part) relating to expectations, whether participants' objectives have been achieved, etc provides a stimulus for small group discussion ('participants will be asked ... to reflect in small groups on what they have gained from the course. For this discussion, the following guidelines will be given: "Consider how successful you have been in achieving the objectives that you set for yourself on the first day of the course. Formulate a plan of action - as teachers and learners - for the future (i.e. how are you going to use what you have learned from the course ?) Each group will be asked to report on its conclusions"' (McGrath, *op.cit.*: 34)). See also Hamp-Lyons 1986.

9. ***structured discussion*** within final session, led by Course Director, on issues arising from end-of-course evaluations

10. ***tutors' meeting***, following the final session, to consider the implications of the findings of the various evaluations for future courses

11. ***Course Director's report***, with recommendations for action

post-course evaluation

12. ***questionnaire*** to participants roughly four months after the end of the course, inviting comments on the effects of the course on their teaching (McGrath 1986 contains an extract from one of the resulting letters)

APPENDIX 6.2

Criteria for the evaluation of TLE

Participant numbers

A self-financing university institution such as IALS needs to generate sufficient income to maintain its staff and the kinds of research and development activity for which it was set up; it is therefore essential that teaching programmes operate at a surplus. (This need not imply that each programme must of necessity make money from its inception, but the incremental income from programmes should at least cover overheads, and ideally do better than this.) One criterion for evaluating the success of a programme will therefore be the money it makes, and one factor in this will be participant numbers. A minimum viable number and a maximum, or 'ceiling', is set for each course. Failure to reach this minimum number has led (*pace* Rudduck 1981 and Mackay 1994) to programmes being cut or courses cancelled and prices being fixed rather than raised.

TLE ran three times in 1982 with a ceiling of 24 and attracted a total of 67 participants (= 93% of its capacity). This can be compared with the two courses in 1981 which attracted a total of 28 participants (no information available on capacity). In 1985, the ceiling for TLE was reduced to 20 per course; between 1988 and 1992, numbers were stable at 50-55 over the three courses, or 83-91% of capacity. On most occasions, there were late cancellations.

This evidence of a sustained increase in the number of participants obviously needs to be seen in the general context of marketplace visibility (and the resources committed to marketing); however, taken together with other criteria, it is a useful index of programme efficiency.

Breadth of appeal

In itself, the number of participants is a crude measure. It has to be seen in relation to fee income, to costs, and - especially in view of vacillations in money markets and the volatile global political situation - to the spread of participants over countries. Breadth of appeal - on which continuing viability might depend - is thus a further criterion. A programme with very broad appeal would not only attract teacher participants from different geographical areas but also, potentially, from public and private sectors and from primary, secondary and tertiary levels.

The table below shows the geographical distribution of TLE participants over the years 1987-1990 inclusive.

NATIONALITY	1987	1988	1989	1990	TOTAL
Italian	13	17	18	15	63
Spanish	10	11	14	17	52
Swiss	3	4	4	5	16
Greek	0	11	2	2	15
French	6	4	3	1	14

Japanese	6	1	4	3	14
German	5	1	1	4	11
Austrian	1	0	1	1	3
Norwegian	1	1	0	0	2
Mexican	0	1	1	0	2
Hungarian	0	0	2	0	2
Palestinian	0	0	2	0	2
Dutch	0	0	0	2	2
Cuban	0	0	0	2	2
Yugoslav	0	0	0	2	2
Chinese	0	1	0	0	1
Jordanian	0	0	1	0	1
Uruguayan	0	0	1	0	1
Swedish	0	0	1	0	1
TOTAL	**45**	**52**	**55**	**54**	**196**

Several conclusions can be drawn from the table, e.g.:

1. the geographical spread ranged from 8 countries in 1987 to 14 countries in 1989
2. participants from Italy and Spain accounted for more than 50% of all participants
3. there was a steady rise in Spanish participants (NB this might have been higher if it had not been controlled by a system of nationality quotas)
4. there was a steady decline in the number of French participants
5. although the course attracted participants from 19 different countries over this period, the vast majority came from only seven countries.

Unfortunately (and this point is taken up in the conclusion to Chapter 6), course records are incomplete for other years in the period surveyed. It is therefore impossible to ascertain whether, for instance, the apparent trends noted in points 2 and 4 above continued or whether the 'new' nationalities featured in 1989 and 1990 can be seen as evidence for a broadening of appeal.

Information on the types of institution in which participants work is similarly variable, only three of the course reports for 1982-1992 giving any quantitative indication of the breakdown of participants across institution types. The course reports for 1986, 1990 and 1992 show that although the majority of teachers on any TLE course work in state secondary schools, and it is for this group that the course was primarily designed, TLE has also attracted teachers from primary schools (5 in 1992), polytechnics and universities, and from the private sector, including those who exist on private lessons (1 in 1990). There have also been teacher trainees (2 in 1986, 6 in 1990), teachers of other languages and teachers of other subjects who wanted to teach English (2 in 1990, 1 in 1992). Each of these exceptional cases, needless to say, was admitted only after a certain amount of negotiation.

Participant satisfaction

The above criteria represent what might be seen as a managerial view of evaluation. If one were to take a more participant-centred view, then participant satisfaction would be of primary concern. (There is also a managerial dimension to this, of course: high levels of

participant satisfaction may mean savings for the institution, in that in the short term less time needs to be spent on programme revision and less money on marketing.). Since methods for evaluating participant satisfaction this are the main focus of Chapter 6, the treatment here is limited to its manifestation in (1) personal recommendations and (2) judgements of value-for-money; change in participants is dealt with in the subsequent section.

Personal recommendation

The logical ink between participant satisfaction, personal recommendation and participant numbers is a fairly obvious one. Information on the number of participants enrolling on IALS courses as a result of personal recommendation is collected by means of a marketing questionnaire, which is completed voluntarily. The 1985 course report, which is the only one to include a figure for personal recommendations, attributes 19 of the 65 enrolments (equivalent to 29%).to this category. Since archive records are only stored for three years it is impossible to establish how far this is typical.

Value for money

Another more direct way of measuring participant satisfaction is through judgements of relative value for money.

Participants on TLE were asked to comment on this each year between 1981 (the year prior to the introduction of TLE) and 1985, and the results noted in course reports. In 1981, a five-point rating scale was used. Responses were as follows:

1. very good value	0
2. quite good value	7 (28%)
3. adequate value	4 (16%)
4. not very good value	14 (56%)
5. poor value	0

For purposes of comparison, exactly the same scale was used for TLE in 1982. Responses were very different:

1. very good value	23 (48%)
2. quite good value	22 (46%)
3. adequate value	3 (6%)
4. not very good value	0
5. poor value	0

Although both sets of results are incomplete (1981: 25 out of 28 participants; 1982: 48/67 participants), they do suggest a significantly higher level of satisfaction in 1982, when no respondents opted for an explicitly negative response (as compared to more than half of the participants in 1981).

In itself such a comparison is only valid if we take into account other considerations such as any differential in cost (either planned or the result of exchange rate movements) or any change in timetabled hours. Over the next few years, however, there was a steady increase

in cost, represented by course fees and a reduction in taught hours, yet value for money ratings remained relatively high, as this modified table for 1985 indicates:

1. good value	14 (48%)
2. reasonable value	14 (48%)
3. poor value	1 (3%)

(Figures here relate to only two of the three courses; this question was inadvertently omitted from the questionnaire given to the third group.)

Change in participants

In-course evaluation of change

If a high level of participant satisfaction is one desirable effect, others might be demonstrable changes in, for instance, partiicpants' knowledge, linguistic competence (if they are non-native speakers of the TL), classroom performance, and attitudes (including confidence). Whether these are realistic objectives for a non-specific short course is doubtful. On the one hand, there is the question of whether changes of an other than superficial kind can be stimulated; on the other, whether any such changes can be observed. Within the space of two to four weeks, there is relatively little time for significant changes to take place, and since refresher courses tend to adopt a wide-ranging approach to content (e.g. a different topic each session), it is unlikely that topics will be dealt with in sufficient depth for existing habits (including linguistic habits) to be displaced. A further inhibiting factor is that the pace of such courses normally allows little time for reflection and assimilation (but see Parker and Haworth 1985 for ways of counteracting this problem). Efforts can naturally be made to ascertain to what extent changes appear to take place during a course, but the real question, as with any programme, is whether effects are sustained. Post-course evaluation of effects, an issue broached in Chapter 2, is clearly a particular problem for open-access courses for overseas teachers.

As far as in-course evaluation is concerned, the only evidence is likely to be that derived from self-report or observation, since quantitative data-gathering, through tests, is unlikely to be acceptable on a non-award bearing course to either participants or tutors.

The clearest statement of the changes anticipated can be found in course objectives (or outcomes). The brief description of TLE used in publicity materials (see 6.2.1), which has seen very little modification since 1982 implies, among other things, that the course will 'extend' participants' 'knowledge of modern developments and techniques' and lead to an improvement in their fluency and 'confidence in using English for teaching and other purposes'. Course reports make no explicit reference to the objectives, but it can perhaps be assumed that evaluation of their fulfilment has been inferred from responses to the end-of-course questionnaire, which has consistently elicited participants' views on the value of the methodological component (as a source of new ideas) and the language learning component. As will be clear from he tables below, responses to these questions are easily converted into statistics that can be used for summative purposes.

TLE as source of new ideas (1992)

NEW IDEAS	n = 52
very valuable	26
valuable	24
quite valuable	2
not very valuable	0

Linguistic value of TLE (1992)

LINGUISTIC VALUE (n = 52)	Weeks 1-3 LANGUAGE ACTIVITIES	Week 1 SPOKEN ENGLISH	Week 2 TEXT ANALYSIS	Week 3 LANGUAGE & CULTURE
very valuable	25	24	19	13
valuable	20	16	21	17
quite valuable	5	6	9	13
not very valuable	2	3	1	6

There is also evidence (see 6.2.3.7) that participants:

- are able to evaluate what they have achieved (in relation to the objectives they formulated initially) and, where relevant, why they have not achieved these objectives (Questionnaire results);
- feel better able to handle problems in their teaching contexts (Questionnaire results).

Occasionally, participants mention other effects that could not have been predicted. Thus, one Spanish teacher commented that although she had found the course useful as a source of teaching ideas, what she had most appreciated was the chance to witness "how a teacher can be with a class", a reference to tutors' relationship with the TLE group (CD recollection).

Course reports and records of tutors' meetings suggest that through observation, tutors have noted further changes in individuals:

- understanding of concepts, the metalanguage used to describe them, and the ability to use this metalanguage;
- confidence in their own teaching practices, and in describing these;
- confidence in their own use of English;
- self-awareness, in relation to learning preferences;
- temporary destabilisation, in relation to either linguistic or pedagogic competence (assumed to be positive in the longer term.)

Post-course evaluation of change

The only evidence of post-course effects is in the form of self-report. This may sometimes be strikingly positive (see the extracts in Appendix 6.8 from a letter from an ex-participant), but even if one were able to accept such evidence at face value, it would be insufficient as a basis from which to draw general conclusions, especially since both participants and their teaching contexts are so mixed (Hamp-Lyons 1986).

APPENDIX 6.3

Extracts from Study Diary

22.8.83

Games

Fascinating. Interesting AND useful. I shall certainly use some of the games (most were new to me). Problem: this is meant for enjoyment. How do you build it into a course so that the course doesn't seem a patchwork of moments without any coherence.

29.8.83

Language analysis

Very interesting discussion on how non semantic aspects of the language re-inforce the meaning conveyed by the text (i.e. place of word in sentence; stress, etc...). I don't think I can spend so much time with my students doing the same thing but I should pay more attention to that so that I can draw their attention to it (it might be useful when we study advertisements).

Vocabulary development

Interesting & valuable. I do feel guilty because I don't teach vocab. systematically. But I have so little time ! I tend to think that an extensive vocab. is not a priority, but it's true that my students are starved for words. O.K then, I'll work on that.

31.8.83

Video tapes

Interesting but a bit frustrating - I think more time could be spent on how to use video tapes - it's an aid which I have always found very difficult to use properly and as a result, of course, I don't use it.

APPENDIX 6.4

Integrated Activity

TEACHER'S NOTES &
INSTRUCTIONS to
STUDENTS

FLUENCY: Role Play Discussion

N.B.. This activity should not be used until people have worked together for a week or more (i.e. until they have got to know each other reasonably well).

PROCEDURE:

1. Divide students into small groups (4-5). No group should have fewer than 4 students. The class-size is immaterial.
2. Hand out (a) Instructions (b) list of Problem Situations - at least 2 of each per group. Make it clear that each student should choose a person from the list whom he/she feels most sympathy for or best understands.
3. Allow a few minutes for students to make their choices, checking that no two students in a group have chosen the same person, then get the first student to sit a little apart to prepare h-self for the ordeal ahead.
4. From this point on, the activity should run itself.... If a group finishes early, one student or the teacher may choose another role.
5. If time permits, each group reports back to the rest of the class.

TIMING: 45-60 mins

INSTRUCTIONS

PROBLEM SITUATIONS

1. The situations below might (and do) occur in language classes. Read through them and decide which of the students you can sympathise with/identify with/argue for most easily.
2. Tell the rest of your group which student you are prepared to defend. In the discussion that will follow later, you will be that student.
3. Decide as a group on an order for the discussion of the problems.
4. The student who is to argue his/her point of view: you should now sit a little apart from the rest of the group and prepare what you are going to say to them. Remember that they are not fellow students: they are teachers.

   The rest of the group: you are teachers at a staff meeting. Choose a chairman (a different chairman, ideally, for the discussion of each problem situation). You must discuss the problem and decide on a course of action/what to say to the student. You may decide in certain situations that it would be preferable if one member of the group spoke privately to the student. In this case, the person chosen would have to report back to the rest of the group.

   When you have decided what to do/say, call the student over...

IMcG
IALS
1982

1. Student A talks a great deal in class, and constantly interrupts the teacher to ask questions.
2. Student B is very quiet. He/she speaks only when forced to, and then seems afraid of making mistakes.
3. Student C thinks his/her English is better than it really is, and takes no notice of corrections.
4. Student D persists in speaking his/her own language to a neighbour. When teachers make them sit in different parts of the room, they talk across other students.
5. Student E is older than the other students and finds it very difficult to work in the way that he/she is expected to; he/she does not mix socially with the rest of the group.
6. Student F thinks that the atmosphere in class is too light-hearted: he/she came to England to work hard, and feels that teachers should be stricter.
7. Student G is always late for the first class of the day.
8. Student H thinks that classes which involve a lot of discussion are a waste of time. He/she wants to learn grammar.
9. Student I appears to dislike Student Z, and when they are asked to work together as a pair or in a group they refuse to speak to each other.
10. Student J feels that individual work can be done out of class and that teachers should use class time for teaching.
11. Student K works well in most classes, but whispers and fidgets during lectures.
12. Student L thinks that there is no point in talking to other students because they make mistakes.
13. Student M has a lot of difficulty understanding authentic taped material, and keeps asking teachers to give him/her the tapescript.
14. Student N has complained because in his/her view teachers do not correct students enough.

IMcG
IALS
1982

APPENDIX 6.5

Example of Forum questions

1. How do you/could you/would you deal with the problem of mixed-ability classes ?
2. Do you have any comment on the argument that groupwork is likely to be unsuccessful because pupils will speak their own language if the teacher is not listening.
3. What suggestions can you make to the teacher with discipline problems ?
4. Would you agree that many of the activities used in classrooms test rather than teach ?
5. We have a responsibility to our pupils and our English-teacher colleagues. Should we therefore concentrate on getting through the prescribed textbook(s) ?
6. Given very limited time for the teaching of English, what can we do to encourage learning out of class-time ?

APPENDIX 6.6

Extract from pro-forma used to elicit participant language-learning objectives

Step One Use the following scale and put a number in the **Level** box to indicate your present ability in the sub-categories listed:

Scale	1	2	3	4
Interpretation	**Fair**	**Satisfactory**	**Good**	**Very Good**

Language Area			Level
Speaking	pronunciation		
	stress/intonation		
	fluency		
Listening to	conversations		
	TV/radio		
	a talk		
Reading	practical information		
	newspapers		
	recreational *(e.g. novels, magazines)*		
	professional/study		
Writing	notes/messages		
	letters	personal	
		business	
	professional/study		
Grammar	correctness		
Vocabulary	range		
	appropriateness		

Step Two Consider which three of these sub-categories you can **realistically** hope to improve within the space of three weeks. Put a circle around them. These are your priority objectives.

Step Three **An Action Plan - How to Achieve Your Objectives**
Transfer the 3 priority objectives you selected above to the box below. Then put a cross (X) in each of the **Means** boxes which may help you to achieve these objectives.

	Means to Achieve Objectives					
	TLE	**Non-TLE Means**				
Priority Objectives	Class Practice	Self Access	Personal Contacts outside IALS	Cinema, TV Radio etc.	Newspapers Books etc.	Others ... ... (specify)
1						
2						
3						

Step Four **Evaluation**
During your individual tutorials we will discuss this plan, your progress in achieving your objectives and consider how you can continue this progress after the course.

TLE.SELFEVAL

APPENDIX 6.7

End-of-course questionnaire

Teaching and Learning English

Section One: Please complete BEFORE Friday

The Right Balance

What do you feel overall about the balance between theory and practice in the first two hours of the day? Please tick (✓) one response.

not enough theory	☐
the right balance	☐
too much theory	☐

New Ideas

How valuable have you found the activities from 9.00am to 10.00am (in Session 1) and 11.30pm to 12.45pm (during Session 2) in giving you ideas for the handling/exploitation of material?
Please tick (✓) a response.

very valuable	☐
valuable	☐
quite valuable	☐
not very valuable	☐

Linguistic Value

How valuable have you found the activities from 9.00am to 10.00am (in Session 1) and during Session 2 (11.30am to 12.45pm) in providing opportunities for you to practise your English and extend your knowledge of the language/culture? Please tick (✓) one response.

	Weeks 1, 2 & 3 **Language Activities**	Week 1 **Spoken English**	Week 2 **Text Study**	Week 3 **Language & Culture**
very valuable	☐	☐	☐	☐
valuable	☐	☐	☐	☐
quite valuable	☐	☐	☐	☐
not very valuable	☐	☐	☐	☐

Pedagogic Value

Put a plus sign (+) next to the **THREE** Language Teaching topics which you found most valuable to you as a teacher (**maximum: three +'s in total**).

If you feel any session should **not** have been included in the course, put a minus sign (-) next to it.

Week 1	**Week 2**	**Week 3**
☐ Language Learning (Day 1)	☐ Teaching Listening	☐ Mixed Level Teaching
☐ Language Teaching (Day 1)	☐ Teaching Grammar	☐ Language Testing
☐ The Communic. Approach	☐ Teaching Writing	☐ Developing Learner Indep.
☐ Managing Spoken Interaction	☐ Teaching Reading	☐ Lesson & Course Planning
☐ Spoken Eng. & Error Corr.	☐ Skills Integration	
☐ Drama in the Class	☐ The Communic. Classroom	
☐ Priorities in Spoken English	☐ Grammar Games	
	☐ Using Video in ELT	

Section Two: Please DO NOT complete before Friday

Was the course different from what you expected? If so, how? **Yes** ☐ **No** ☐

Have you achieved the language-learning objectives that you set yourself at the beginning of the course? If not, why not? **Yes** ☐ **No** ☐

Do you feel better able to cope with major problems in your classes? **Yes** ☐ **No** ☐

Do you have any suggestions for improving the course or our resources?

What did you like about the course?

Resources

How valuable have you found the available resources? Please tick one response for each resource.

	Resource Centre	Study Room	Language Laboratory
very useful	☐	☐	☐
useful	☐	☐	☐
not useful	☐	☐	☐
never used them	☐	☐	☐

Self-Access

Did you use any self-access material? Yes ☐ No ☐

If yes, which area(s) did you focus on?

language improvement ☐

materials design ☐

reading references ☐

videos on classroom teaching ☐

How useful did you find the material?

very useful ☐

useful ☐

quite useful ☐

not useful ☐

Any comments: ______________________________

____

____

____

____

____

____

APPENDIX 6.8

Extract from post-course letter

January 5th 1986

... However strange it might sound, I'd like to start with the conclusion I arrived at after these last three months and that is: get rid of the coursebook and design your own syllabus according to the learner's demand. It may sound a bit drastic but I never felt so free and so happy as in September 1985. While revising, I chose some general enough topics so as to motivate my learners into thinking in terms of their own needs and therefore elicit from them the appropriate forms or else demand from the teacher and/or search for the necessary items in the dictionary, which I now use freely in the classroom. Thus the learners were generating their own language, the new linguistic items generated new activities, which in ever-widening circle generated new needs, new language and so on and so forth. In such a way I could have designed a new syllabus step by step; they all spoke, listened, read and wrote, even the weaker students who never used to open their mouths.

Unfortunately, after a while I had to use the coursebook. Had to ! Why ? I didn't "have the guts" to go on as I had started (too much pressure). Yet it had taught me that mixed ability classes were not an impediment, I'd even say that one should have mixed ability classes (whereas, up to last year, I had never been able to make up my mind on that particular issue). Students teach each other, the weaker students seem to be more receptive, less shy when taught by their peers. That is the result of group work, role play, or simulation - I have done a lot of group work, although they don't always speak English and although the pronunciation seems to suffer a bit, they feel highly motivated and usually come up with some tangible result by the end of the period ...

As you can see, my role as a teacher has changed to some extent, so has the classroom management. It has changed my students and I have changed too in relation to my work; less bored, eager to try new "tricks"....

Everything is not so rosy as it may sound in this letter but on the whole it has become much more interesting. Thank you for the great big booster you gave me....

APPENDIX 7.1

INSET for language teachers in the Lothian Region of Scotland

1. DATA SOURCE

The information contained in this Appendix derives from two interviews conducted in late 1992 with Peter Wheeldon (PW), then Languages Adviser for Lothian Region. Quotations are from an edited transcript of the interviews. The section has been checked for accuracy by PW.

2. THE INSET POPULATION

As Languages Adviser in Lothian, PW has responsibility for the following:

- state secondary schools: 48 (approximately 250 teachers of languages: 210 permanent full-time and 40 temporary part-time);

- primary schools: 240 (approximately 30 staff actively engaged in Primary Languages Project, but many more dealing with language awareness or communications systems as part of project-based programmes).

Although the Adviser has no responsibility in respect of the private sector, it is the policy of the Education Department to give teachers in private schools access to the in-service opportunities organised for their state-sector colleagues. Although a fee is payable, this option is frequently taken up by teachers in the private sector, suggesting that some value is placed on this provision.

3. FUNDING AND CONTROL OF BUDGET

INSET funding comes out of a 'folio' for the Advisory Service. This pays for all forms of INSET provision, from the direct costs of mounting one-day meetings and longer courses to the expenses of sending teachers to do courses elsewhere. Other sources of funding are the Central Bureau for Educational Visits and Exchanges (particularly for overseas visits) and Lingua (e.g. for primary teacher exchange programmes). A recent development has been the devolving of responsibility for small-scale funding (items of less than £100) to the schools themselves.

The budget for INSET (which also covers the INSET needs of Advisers) is managed by the In-service Department. Since it is needs-based rather than strictly apportioned to specific areas of the curriculum, a flexible response is possible to changing circumstances, such as the move towards compulsory languages from Secondary Grade 3 or the introduction of foreign languages in the primary schools.

4. ORGANISATION OF INSET

4.1 Forms of provision

The INSET programme devised by the Adviser has taken two basic forms:

1. centrally-mounted one-day meetings and 'twilight' courses (consisting of a series of once-a-week meetings after school for anything between ten weeks and eighteen months);
2. meetings of 'neighbourhood groups' of teachers (i.e. schools in the same geographical area) within the school day in time that has been set aside for professional development. For these meetings, the Adviser travels to the neighbourhood group rather than vice-versa. Teachers are advised of these through a booklet circulated to schools.

Significantly, perhaps, the twilight courses are relegated to an Addendum within the booklet. Within Lothian, there has been a marked shift over the last few years from (1) to (2). Some of the advantages and disadvantages of the two forms of provision are surveyed below, under 4.4.

4.2 Input to the content of the INSET programme

Ideas for the content of the INSET programme come from three principal sources:

1. national initiatives (e.g. 5-14), where INSET is a response to a clear need for information as well as training and education;
2. expressed interests of teachers, prompted by a menu of possibilities circulated by the Adviser;
3. more *ad hoc* suggestions for follow-up or other events from teachers who have just attended a meeting or course.

4.3 The aims of INSET

Like most other INSET providers (see, e.g.), PW sees INSET in terms of change:

> 'it is making people who are willing themselves to try and become a little different to try things they've never done before. It does require a change in priorities, a change in a little of the structure in the classroom, a change in attitudes, a lot of change which I think is very good: a fearlessness of being wrong, a risk-taking attitude erm and an increasing degree of objectivity ...about evaluating what you do. So ... as far as I'm concerned staff development means changing something - we hope of course ... for the better for the teacher and the learner.'

4.4 Centre-based longer-term courses *vs* school-day meetings

The shift from centre-based to neighbourhood INSET appears to have been motivated largely by practical considerations. For instance, there is no loss of teacher time (and therefore no need for substitution) since the hours spent at a neighbourhood INSET day are timetabled hours which have been 'gained' by teaching a series of extra half hours. Time might also be seen as a factor if travel to the Centre is taken into account.

However, PW expresses some unease regarding the effectiveness of the one-day neighbourhood meetings:

'The word "course" is still being used for many of the things that I would label one-day meetings ... it is an in-service meeting and I'm often doubtful of the value of doing something for one day only. I'm not saying it's useless but I don't think it's the best way to do it ... changes in people's attitudes require a bit of time ... the fact that you've been given time doesn't guarantee they'll change er but I my point is I don't think attitudes change in one day so no change will take place and as far as I'm concerned in in-service if no change takes place between the ears ... then no change will take place within the classroom.'

What can be observed in respect of the Centre-based twilight courses is commitment: as PW points out, to attend an after-school course once a week for ten weeks 'betokens a level of commitment to what is being offered and what they want to get from it. As far as one-day meetings are concerned, we don't know what the level of commitment is ... it isn't put to the test'. Within a longer course, moreover, there is time to observe change, in all concerned, and time for rapport to develop - 'it's richer'.

The one-day events, by contrast, produce less tangible results. Although PW gives follow-up material in the form of suggested readings, he has no way of knowing whether any use is made of these. Nor does he have any way other than through questionnaires of ascertaining the effects of the day itself: 'I don't know how I'm doing with them and there's not a lot of chances through the system to find out'.

Other practical disadvantages of neighbourhood-based INSET include the closed nature of these groups, which are bound together by geographical proximity rather than perceived common interests; and the duplication of effort on the part of the Advisor, who now travels to the neighbourhood groups, sometimes even to single schools, to repeat the same 'course'.

5. EVALUATION OF INSET PROVISION

5.1 Forms of evaluation

Evaluation of L2 INSET takes two forms:

1. attitude-revealing activities adapted from a German model;
2. a short standard questionnaire specially devised by the Adviser, which includes multiple-choice questions to elicit level of satisfaction with the event and open-ended questions to prompt both suggestions for improvement and other wishes (see Appendix 7.2).

(1) enables the Adviser to assess any (effects of) attitude-change in subsequent school visits; (2) enables him to get a feel for participant response.

5.1.1 Attitude-revealing activities

'Krumm cards'

The Krumm cards (named after their originator) contain statements which express beliefs, e.g. 'Boys taught by women teachers do better than boys taught by men teachers' (from a set of cards on the topic of motivation). The first part of the activity involves participants taking a card at random and attempting (in the role of devil's advocate, if necessary) to get others to agree to the

statement contained on the card. Agreement is indicated by a signature on the card (pseudonyms can be used). Once a card bears two signatures, it can be put into a container for signed cards. Cards which attract no signatures are placed in a separate container. The activity continues until all the cards have been exhausted. At this point, the cards are pinned up, under category headings, on a large noticeboard and a pulling-together discussion follows on the patterns of groupings and any apparent contradictions across these.

The key phase of the activity is, of course, the ensuing discussion, which throws up issues which relate directly to the content of the course and the group experience. From the Adviser's point of view, such a discussion can be immensely revealing in terms of group awareness and the extent to which this has been influenced by the course.

Priority discussion

The second instrument for monitoring course effects also makes use of statements. In this case, ten statements are contained on a single sheet of paper and participants are asked to rank these from 1 to 10 in accordance with their own beliefs. These might relate to the characteristics of the Good Languages Teacher, for example. The sheets are completed individually, and the results are then tallied for the whole group, using an OHP. A group ordering can be calculated by simple arithmetical procedures and discussed. Additionally, or alternatively, the group order can be compared with the individual's own ordering. Any differences, and the basis for the individual's decisions, can be examined in one-to-one consultations with the Adviser if the framework of the course allows for this. (At one recent course, 8 out of 13 participants took up the invitation to talk through their choices - and how these were reflected in their classrooms.) At the individual level, this gives the Adviser the opportunity to pinpoint individuals who are likely to need further support: 'That's how I know where to visit'.

There are at least three potential limitations to the use of such activities as evaluation measures. First, what they appear to reveal is current attitude or level of awareness rather than any change in this respect. Second, teachers may tell the Adviser what they think he wants to hear (this implies a certain kind of teacher-Adviser relationship). If they can do this, PW says, then the sheet is badly designed, since all the statements must sound positive and have at least some discernible merit. The third possible drawback is that teachers may be well able to verbalise ideas which are not reflected in their practice. The Adviser's visits to schools allow him to check whether this is the case. At the end of a course, he makes up packages containing materials the teachers have produced or requested. Taking these round schools affords a natural opportunity for seeing teachers in their working environment and asking what they have been able to do since the course. Tangible effects are what is important: 'are they espousing the philosophy/beliefs behind these ideas which will materialise themselves in the classroom?'.

5.1.2 Questionnaire

The standard questionnaire developed out of preliminary reading (of applied psychology books) and the experience of designing a questionnaire for use with school-age learners (the GLAFLL questionnaire). The aim was to produce an instrument which would be perceived as friendly and would prompt honest comment.

Within the context of the twilight courses, more than one questionnaire might be used, especially if the course content is organised in more or less self-contained units. For one-day meetings, the questionnaire is used at the end of the day (see Appendix 7.2 for examples). The drawbacks of the

latter procedure (e.g. hasty ill-considered completion) are recognised but the alternative would be a low rate of possibly unrepresentative returns.

> 'If you give them a week or two to reflect you've got in-service evaluation of the people that wanted to tell you something, the people at the extremes, the people who are affected enough for long enough and it was the minority. Now that might say something in itself.'

The difficulty with this position is that it remains speculative. No response to a post-meeting questionnaire does not necessarily mean that nothing is taking place; nor, for the matter, can one be certain from a positive response that something is happening as a result of the meeting.

5.2 Evaluating the secondary effects - on the learners

If an Adviser can get into classrooms, it should not be too difficult to observe the effects of INSET on teachers. If they are doing things they were not doing before, INSET has had some effect (this assumes one knows what they were doing before).

However, the ultimate aim of most INSET is to produce effects in learners, and this, as PW observes, can be difficult to evaluate:

> 'what we can't, say, evaluate, is is it better for the kids and them that they are doing it ? We can say that we've changed them, which is much less arrogant than saying we've developed them.... What I'd liketo know is is what he or she is doing now considered by the teacher and the learners to be better, more effective, more motivating, more something positive. And [the evidence for] that is only verbal report by the teacher.'

However, the Adviser does have a way of assessing the validity of these verbal reports when he is invited into a school by a teacher to try out the ideas that have been discussed on a course.

> 'the valuable things that we get from doing this are the ones that are least measurable: children happy, involved, making working noise ...I can't show you in numbers. I can't reduce the happiness in the Primary Project of the Primary 7 children running happily to the secondary school to continue with their French or German ... I can't measure adequately for other observers the extent to which these children dote on the languages teachers in their first month at school... You don't measure that sort of thing - you just write a report and say "This was so"'.

The problem with this kind of evidence of effectiveness, as PW recognises, is that because it is not measurable it is not objective, in the normal sense of the word:

> 'to me that's evidence, but it's not ... vulnerable evidence, i.e. I can't lay it out on the table to have it criticised.... Non-vulnerable evidence is therefore less valuable evidence because ... we weren't there: we can't see if any of the kids were miserable deep down; we just have to see things through Peter's ... eyes.... The thing was evanescent, transient ... an event not a fact... Your measuring-stick is highly educated and informed subjectivity....'

He concludes:

> 'I find evidence-collecting, where it really counts and matters and where it's really valuable, I find it exceedingly difficult.'

5.3 Obstacles to further development

Further development of evaluation procedures is unlikely unless two modes of support are made available to the Adviser: feedback on the procedures currently being used and time in the form of colleagues who could share the advisory role. In itself, the first would not necessarily involve any great cost; but it might well argue for the second.

6. TWILIGHT COURSES FOR PROBATIONERS AND RETURNERS: 1991-93

6.1 Introduction

Due to falling school rolls and school closures, very few new language teachers were taken on by the Region for several years prior to the academic year 1991-92. During this period, such support as was provided by the Advisory Service was individual and largely reactive. With the introduction of compulsory language-learning beyond S2 the situation changed. More new teachers were required and the provision of in-service courses to meet their needs became both necessary and feasible.

6.2 The courses in 1991-92 and 1992-93: organisation and content

In each of the academic years 1991-92 and 1992-93 the Region ran a twilight course for a mixed group of probationers (trained teachers in their first or second year of teaching) and returners (mainly women with children, who had been out of teaching for some years). Some of the latter were teaching part-time. The courses were voluntary. The first course, which started in January 1992, attracted 18 participants; the second, which started in September 1992, 12. In each case, probationers and returners were more or less evenly represented.

Both courses were organised on a once-a-week basis (16.30-18.00), the first three to four sessions being scheduled and subsequent sessions being arranged on demand. In the event, both courses continued for 11 weeks.

Content was determined by the participants. The Adviser, working with three colleagues, offered a 'menu' of content options - "basically lifelines" (e.g. updating on exam requirements and assessment criteria, methodological developments, motivating the least able). Participants were asked to decide on an individual basis which of these topics interested them most and which other topics they would also like to explore. Then, after some negotiation within the group as to priorities and sequencing, subgroups were formed to look at the topics in question. The subgroups worked autonomously, each deciding how long to spend on the topic(s) they had selected. One result of this was that some groups and participants continued for longer than others.

6.3 Underlying principles

Judging from the Adviser's description of the courses (see above) and further comments (interviews of March 22 and March 29, 1993), the organisation of the courses seems to have been underpinned by certain participant-centred principles:

1. Participants would themselves determine the length of the course (initially, they were only asked to commit themselves to three or four weeks because it was felt that there might be reluctance to any longer-term commitment).

2. The syllabus should be negotiated but essentially participant-driven. Tutors should respond flexibly to participant needs.

3. A distinction would be made between 'wants' ("I want to know how to use these materials - and I want copies of the materials") and 'needs' (as defined by the Adviser, an understanding of the principles on which a particular type of activity or set of materials is based, and the skills to make the materials). In other words, participants would be encouraged by course processes to move from a dependency-orientation (i.e. wanting ready-made answers) to a more self-directed mode of thought (i.e. "What do I need to know to be able to do this for myself?"). One element in this was the use when appropriate of short between-session reading tasks which, it was hoped, would both inform and prompt further reading.

4. The presentation of new, theoretical ideas through reading and discussion would make the course "more cognitively demanding than just going into a shop and buying things" (an extension of the menu metaphor). Participants would be stimulated to consider the relevance and applicability of these ideas for their own situations. (See also Evaluation, below.)

6.4 Evaluation

One assumption underlying the mixed composition of the participant group was that despite differences in age and experience (many of the returners had substantial experience), all would have "beginner-y" needs with regard to certain of the topics suggested. The validity of this assumption was checked at the first meeting, when participants were asked to state their preferences (and other interests). This process constituted a form of needs analysis which took into account the perceptions of both participants and trainers.

Since the tutorial team was made up of two College lecturers responsible for supervising teaching practice and two Advisers who were in regular contact both with teachers in schools and the Principal Teachers responsible for supervising the work of new teachers, it seems highly unlikely that a specification of needs based on the tutors' experience would differ much from a more objectively derived set of needs, even if this had been possible. A more direct approach to needs analysis in relation to the probationers, involving the observation of novice language teachers in the schools of the region, would not have been possible prior to the first course since - as stated above - so few new teachers were being employed.

A second assumption was that the mix of participants would be positively beneficial, that they would have "different strengths" and "enrich" each other. Participants' reactions to the mix were not formally evaluated. It was noted, however, that although probationers were initially noticeably more silent than returners, the former proved to know more about certain of the topics discussed, such as pairwork; in other words, the returners' previous learning had to some extent been superseded by more recent developments.

The Region's basic method of collecting feedback on in-service courses is by means of questionnaires to schools. However, the questionnaire is used to gauge teachers' response to the

sufficiency and adequacy of in-service provision in general rather than to evaluate the perceived effectiveness of the programmes arranged by specific advisers. And since the advisers themselves are under no obligation to justify the continuation of their budget allocation by rendering an account of the programmes they offer, it would not be surprising if - given the many demands on his time - an adviser were content with informal and indirect evidence that a programme had met participants' perceived needs. In fact, an additional standard questionnaire has been devised by the Aesthetic and Linguistic Team to evaluate one-day events. As the examples in Appendix 7.2 illustrate, however, these tend to be completed hastily and are relatively uninformative.

No formal written evaluation of the two twilight courses was carried out. In fact, those who participated in the second course actively resisted an attempt to get them to express their comments in writing, arguing that their continued attendance and positive oral comments were evidence enough of their positive attitude. Attendance registers might indeed have offered indirect evidence of the perceived value of the course. Unfortunately, such attendance records as exist are incomplete.

Some feedback on the effects of the course was obtained subsequently in the course of visits to schools. However, this was largely incidental - the visit often having to serve a variety of purposes - and no systematic records were kept.

APPENDIX 7.2

Examples of participant reactions to Lothian Region L2 INSET courses using standard questionnaire

AESTHETIC AND LINGUISTIC TEAM

In-Service

Subject LANGS ***Date*** 6/2/90

EVALUATION QUESTIONNAIRE

Name: *(Optional)*

1 What did you think of today's workshop/meeting?

very useful; (circled)
useful;
alright;
not very useful;
a waste of time?

2 Which part of the workshop/meeting did you find

a) most valuable;

Talk from Mrs Smith of Exam Board

b) least valuable?

Not applicable

3 What aspect which was not included would have made it a better workshop/meeting for you?

Not applicable.

4 Any other comments: ***(continue overleaf if necessary)***

a refreshing and enjoyable day.
Time very well spent.

AESTHETIC AND LINGUISTIC TEAM

In-Service

Subject...Mod. Langs – P.T.s Meeting... ***Date***...06/02/90...

EVALUATION QUESTIONNAIRE

Name: ***(Optional)***

1 What did you think of today's workshop/meeting?

very useful; (circled)
useful;
alright;
not very useful;
a waste of time?

2 Which part of the workshop/meeting did you find

a) most valuable; All

b) least valuable?

3 What aspect which was not included would have made it a better workshop/meeting for you?

4 Any other comments: *(continue overleaf if necessary)*

AESTHETIC AND LINGUISTIC TEAM

In-Service

Subject..... Modern Languages .... ***Date***..... 6/2/90 ....

EVALUATION QUESTIONNAIRE

Name: ***(Optional)***

1 What did you think of today's workshop/meeting?

very useful; (circled)
useful;
alright;
not very useful;
a waste of time?

2 Which part of the workshop/meeting did you find

a) most valuable;

alles!

b) least valuable?

nichts!

3 What aspect which was not included would have made it a better workshop/meeting for you?

nichts!

4 Any other comments: ***(continue overleaf if necessary)***

ausgezeichnet / wunderbar!

AESTHETIC AND LINGUISTIC TEAM

In-Service

***Subject*....Languages................ *Date*...6 Feb 90......**

EVALUATION QUESTIONNAIRE

Name: ***(Optional)*** Anne B. Johnstone

1 What did you think of today's ~~workshop~~/meeting?

very useful; (circled)
useful;
alright;
not very useful;
a waste of time?

2 Which part of the workshop/meeting did you find

a) most valuable;

ALL

b) least valuable?

—

3 What aspect which was not included would have made it a better workshop/meeting for you?

More detailed consideration of Revised H. & Revised CSYS

4 Any other comments: ***(continue overleaf if necessary)***

There is never enough time to deal with all matters causing concern at a time of tremendous changes.

APPENDIX 7.3

Course 93T130

93T130	Returners and Probationers Course	16,23,30 November 1993 7,14 December 1993 18 January 1994 4.15pm - 6.00pm	Darroch EC

COURSE AIMS AND OBJECTIVES

1. Returners - Updating on all recent developments in Modern Languages.
2. Probationers - Indentification of needs and their fulfilment.

COURSE CONTENT

Content will be decided by the needs expressed by the course participants, initially from the following menu:-

METHODOLOGY
Some principles
Acquisistion and Learning
Communicative approaches
The slower learner-motivation
Pupil-centred approaches
Autonomy (how much?)
Independence training
Group Tasks
Self-access provision
Dictionary Skills
Dictionary Skills in Examinations
Homework

EXAMINATIONS
Standard Grade :
Speaking
Listening
Reading
Writing
Higher:
Reading
Listening
Cloze Test
Speaking Test
Writing
Modules
Assessment

MATERIALS
Differentiated
Multiskill
Problem-solving
Games

MANAGING
Self
Foreign Assistant
Classroom

TARGET AUDIENCE

First and second year probation teachers and recent returners to Modern Languages teaching (part-time or full-time).

METHOD OF EVALUATION

Questionnaire

NUMBER OF PLACES ON COURSE

25 Maximum

APPENDIX 7.4

Keeping a course journal

The value of keeping a journal

aiding retention

When I introduced the idea of keeping a course journal, I made the point that it can be a useful learning resource: that the act of reviewing your notes can aid retention of course content.

stimulating ideas

Subsequent reflection on your notes in writing can be even more useful. Writing, it has been said, 'stimulates and shapes ideas'.

facilitating integration

Sometimes you may find that there are flashes of understanding, a recognition of how a new idea or piece of knowledge relates to previous knowledge or experience, or how it can be put to use in the classroom. The journal, in this case, can be said to facilitate integration of newly acquired knowledge.

prompting reflection on teaching and learning

Other teachers have used a journal as an outlet for their reactions as a 'learner' to activities, procedures, other participants. These comments, implicitly or explicitly evaluative, can in turn prompt reflection on the reactions of pupils in their own classes; they may also be self-evaluative.

feedback to course tutor

If a course tutor has access to the journals - sometimes they form the basis for a dialogue between course participant and tutor - they can also help the tutor to understand a participant's problems, needs and wishes, and give advice or make adjustments to the course as appropriate. In this case, the procedure followed is rather different, but your evaluative comments will be noted and will benefit participants on the next course (see *What use will be made of the journals?*, below).

Who will read the journal ?

You and me. No one else.

What use will be made of the journals ?

I will ask to borrow your journal at the end of Week 5 (December 14) and return it at the beginning of the next session (January 18). In the meantime, I shall be looking through the journals for anything of interest, but particularly for patterns (e.g. in reactions and concerns), and for any

mismatches between the journals kept by participants and tutors and my own observation notes. The results will provide input to an evaluation discussion during the final meeting. During this discussion, I will not quote directly from your journal or make any references which will allow you to be identified.

As soon as possible after the meeting, I will produce a report for Peter. This will draw on the discussion, the journals and my own perceptions. If at that stage I wish to quote from your journal I will ask for your permission. Any quotations can be anonymous, of course.

Thank you for your cooperation.

Ian McGrath

APPENDIX 7.5

Analysis of participant journals

Following a brief indication of the organisation and content of each session, participants' comments are categorised. It should be remembered that of the 19 participant journals examined only ten contained comments, and that these journals were not all completed on a session-by-session basis.

The following symbols are used: + indicates an overtly positive comment, - an explicitly negative comment, and ? an expression of doubt. 'Value' conflates comments on informativity, relevance, practicality and generalisability; 'interest' covers novelty (e.g. practical ideas) and what was felt to have been gained from others' experience.

Session 1

PW, CS. Brief introduction, then division into two roughly even groups according to participant choices: PW (Methodology - an introduction to communicative language teaching), CS (Standard Grade).

Participants commented on:

- value of session (5+/1-)
- reassurance (2+)
- learning from other participants (2+)
- participant-led approach (2+)
- positive expectations (1+)
- PW's sense of humour (1+)
- comfortable, friendly group (1+)
- time-keeping at end of session (1+)
- needs of different groups recognised (2+/4?)
- fulfilment of needs within 6 weeks (1?)
- organisation/use of menu/group formation (5-)
- no opportunity for participant self-introduction (3-)
- materials (2-)
- jargon (1-)
- domination by practising teachers (1-)

Participants also express anxiety concerning their ability to respond to the (changed) demands of the classroom. One person describes the changes as 'daunting', another is concerrned about the assessment of speaking, and differentiation. Several admit to feeling rather overwhelmed by the first session ('confusion', 'disorganised', 'felt swamped').

Session 2

PW, CS. Continuation of session 1; same groups, same topics, same tutors.

- value (4+/2-)
- interest (2+)
- support materials (1+/2-)

Session 3

PW, CS, AMcD. Approaches to S1. 2 groups (CS, AMcD); groups change rooms at half-way point.

- value (4+/1-)
- interest (5+)
- reassurance (1+)
- questions of returners (teaching)
- useful (1+)
- unable to continue with topic started in Session 2 (2-)
- organisation of group (1-)
- support materials (1-)

Two participants comment on the inclusion of a session on the Learning Support Service as session 4: one is doubtful about the value of this on a six-session course; the other thinks it a good idea.

Session 4

Learning Support Services (AB).

- value (4+/1-)
- interest (4+)
- organisation (1+)
- support materials (1+/1?)
- questions of returners (teaching) useful (1+)
- recognition of PW's efforts as course organiser (1+)
- no examples of differentiation (1-)
- jargon (1-)
- presentation of (new) options for next session (1-)
- relevance to returners not yet teaching (?)

Session 5

Methodology (PW) and Standard Grade (CS) offered to those who had not selected these in Session 1.

- value (4+)
- liking for small-group discussion within group (1+)
- daunted (2-)

APPENDIX 7.6

Questionnaire

Name ....

[ ] Probationer
[ ] Returner (currently teaching)
[ ] Returner

Please circle the appropriate number to indicate whether you (strongly) agree or disagree with each statement.

	strongly AGREE				strongly DISAGREE
1. objectives of the course were clear	1	2	3	4	5
2. good idea to let participants determine course content	1	2	3	4	5
3. good idea to have menu of topics	1	2	3	4	5
4. way topics were selected from menu was well organised	1	2	3	4	5
5. good idea to break into two groups	1	2	3	4	5
6. way groups were formed was well organised	1	2	3	4	5
7. course content has been relevant to my needs	1	2	3	4	5
8. topics have been systematically ordered	1	2	3	4	5
9. session on LSS useful	1	2	3	4	5
10. liked meeting each week in the staffroom over coffee	1	2	3	4	5
11. sessions started punctually	1	2	3	4	5
12. time during sessions has been well used	1	2	3	4	5
13. opportunities to learn from experienced tutors	1	2	3	4	5
14. opportunities to learn from other participants	1	2	3	4	5
15. opportunities to discuss matters of mutual concern	1	2	3	4	5
16. feel I have got to know other participants	1	2	3	4	5
17. sessions finished punctually	1	2	3	4	5
18. good use made of support materials (handouts, etc)	1	2	3	4	5
19. now feel more confident	1	2	3	4	5

And complete the following statement:

20. I would have liked ...

APPENDIX 8.1

Letter 1

[LOGO]

[RETYPED]

December 18, 1994

Dear

This comes with the hope that all has been well since your return home and with a festive greeting:

> May your Christmas be merry
> and your New Year very ...!

Between the sleighing and the schnappsing (or the local equivalent) you might be thinking about your New Year's Resolutions. Which might lead you to think about previous good resolutions, including those that you formulated, with due care and consideration, on the last morning of the course in Edinburgh. Just in case you've forgotten what they were, but still vaguely remember being in Edinburgh (oh yes, sitting out in the garden of the Pear Tree on a balmy (?) October evening, the country dancing, the ghost walk, the pub quiz), I'm enclosing a copy of your ACTION PLAN, together with some light reading for you, your friends or your students [brochures].

People often leave courses feeling rather euphoric, and after the 24-hour-a-day course that you created for yourselves, that would have been entirely appropriate. However, the big question is whether, after this interval, you feel that the course has had any professional impact - that is, has made an appreciable difference to what you know or what you do or how you think, This is the big or broad question; the narrower one is whether you have made any progress as far as your Action Plan is concerned.

You may find it useful to spend a little time at this stage reflecting on these questions (echoes of the session on Self-Evaluation); if you could reflect on paper - and send that paper to me, that would be very helpful for my purposes in my institutional role (course evaluation and self-evaluation). I think I told you during my self-introduction at the beginning of the course that I'm doing a doctorate in the area of Evaluation (of in-service courses for teachers) and action plans are one of the evaluation measures that I deal with in the thesis, so any comments that you might wish to add on to the value or otherwise of action plans would be doubly appreciated!

I hope to hear from you - and, indeed, to see you again at some point, perhaps on your territory next time.

With best wishes,

Ian McGrath
Development Coordinator
Teacher Evaluation

APPENDIX 8.2

Letter 2

[LOGO]
[RETYPED]

February 28, 1995

Dear

Please forgive another 'standard' letter.

I thought you might be interested in a conversation I had last week with one of my colleagues. I said that I feel a little disappointed that so far I've only had one response to my letter to your group asking about progress in relation to your action plans, and added that my timing had probably been wrong, that you were all too busy enjoying yourselves to pay much notice when my letter arrived - or too busy preparing for the new term.

He suggested that there might be another, quite different reason, that you all feel rather embarrassed at having little or nothing to report. Embarrassed because you feel you should have done more, or embarrassed on my account, because I might think that the course has been unsuccessful if nothing happens as a result.

He went on to point out that you were here for only two weeks. This was a very short time in which to achieve anything significant and that, in any case, the effects of educational (as opposed to training courses) may not manifest themselves in concrete terms for some time, if at all. In other words, perhaps I was expecting too much. He also implied that two weeks was too short a time in which to establish the kind of relationship that induces people to respond to requests of this kind out of a sense of loyalty or responsibility. I don't accept this point, but I have scant evidence with which to refute it.

I mentioned that I have had one response and it bears out to some extent my colleague's speculations. I'm sure the writer won't mind if a quote a few sections from the letter:

> As I know your letter would come - I hadn't expected it as soon as this, though - I had already tried to watch myself a bit more closely in my teaching behaviour, and was somewhat disappointed.
>
> ... by now I hope I have understood that I can't become an entirely new teacher within a fortnight, that this is a slow process which will take a long time, which can only be done in small steps and will perhaps never be fully completed. What I have already started to do is trying to provide phases in exercises which are less teacher-centred, and I am trying to use drama-techniques in the teaching of literature.
>
> ... upon reflection I think I profited from my course at your institute more than I thought and in a different way, being made more prepared to reflect what I'm doing and how I could change my teaching habits.

As you will have gathered, I would very much like to hear from you, to confirm or disconfirm the ideas expressed in the earlier part of this letter. My impression was that you were an unusually well motivated group. What actually happened when you got back home?

Very best wishes,

Ian McGrath

APPENDIX 8.3

Examples of Action Plans and Reports on Action

LE

MY ACTION PLAN

I will definitely ...

1. 'cascade' the information/ material/ideas to my colleagues
2. improve my teaching mixed levels methods
3. teach 'language through literature'

(

I also hope to ...

4. pass on to students the fun I had learning the folkdances
5. do self-evaluation more thoroughly
6. to develop learners' independence

NAME Edda Fricke DATE October 21, 1994

ADDRESS

Vor den Hegen 6
21521 Wohltorf

LE

March 13,1995

Dear Ian,

Time runs fast and I am sorry you had to send a second letter. Nevertheless I enjoyed reading your reminding note and realized again what a good psychologist you are. Now that I know I am not the only one being delayed in response I don't feel so bad.

My evaluation:
I often think back to the course and everything around my stay. Now and then I glimpse through my notes but very superficially. And then I get disappointed for the following reason: There I got this wonderful opportunity for this course and the state spent a lot of moneyon it. Then back home I am so obsorbed in everyday routine and tasks that there is no time left to study my notes and books I bought thoroughly so my students and colleaques and I could profit from it!

- Two appointments were cancelled when I meant to "cascade" to my colleques what I had learned.
- I haven't improved my teaching mixed levels methods nor even tried to.
- I taught 'language through literature' and both my students and I had great fun. They were even willing to exhibit their "pieces of art". So we will do it again.
- I haven't taught any Scottish folkdances yet.
- I haven't done any thorough self-evaluation though it's in the back of my head and I seem to realize the mistakes I do over and over again and to look at them more objectively.
- I haven't worked with "Learning to learn English" by Ellis/Sinclair. It's in my mind to develop my pupils' independence and instead of being a prompter to let them find out, correct and experiment themselves.

So now you know. The course hasn't been unsuccessful nor have I been too lazy. Teaching obligations and duties leave no or little time for extras. In my opinion it's a general deplorable state of affairs. But that won't refrain me from attending courses again, as they do make an appreciable difference to me before and after the course.

It would be interesting to know what you find out in your survey on evaluation and action plans. May be you can tell about it at some time here.

With very best wishes,

LB

MY ACTION PLAN

I will definitely ...

1. give more thought to ways of developing student/probationer independence in my classes/seminars.
2. use poetry in classes as a means of communication & motivation
3. get in touch with interested colleagues to work out strategies towards learner-centred activities.

I also hope to ...

4.

5.

6.

NAME Kurt W. Beckmann DATE October 21, 1994

ADDRESS

LB

Siebenschön 41
22529 Hamburg

26 March 1995

ear Ian,

ou're absolutely right, it's about time we gave you some sort of feed-ack. But, you know these things sometimes take an embarrassingly long me to proceed from the in- to the out-tray.

o this is what I think I can say so far: Even before the course at your stitute, terms such as *student-oriented approach* and *learner independen-* e ("autonomy" is the in-word) were the rage in this part of the world. I ad heard of these and read about them, even used them in my own clas-es. Owing to your course, however, it has become quite clear to me that ney must play an even more dominant role than I had thought. (By the ay, I believe that this is also to do with the student of the 90s, who is many ways different from his predecessors in the 70s and 80s, let alone s fossils.) As a result, too, I have been constantly on the look-out for ays and means to bring about more learner independence.

case in point may be corrections of written work. I now tend to mark hat I find objectionable without any further comment. On handing back ne marked papers I ask the students to use their reference books (which ften look immaculate!) and dictionaries and find out from them the cor-ect form/spelling/phrasing etc. Of course, I walk around the class and elp them individually to make sure they understand this time. The other ay, a girl told me that she never knows the difference between the pre-ent perfect and the past perfect. She would never learn it if I just wrote ne correct form in the margin. She would just nod and say "Oh, yes." but ould not feel the need to think it over. By throwing them back on their wn resources we make students realize how important these skills are nd they will, hopefully, make use of them in further written work. Work echniques therefore have become an essential in my teaching them how to arn.

Then working with probationary teachers in my seminars, too, I have given rominence to learner independence by making them draw on their consi-erabie experience and having them discuss things amongst themselves be-ore I feed in additional material and ideas. We find this very satisfactory.

o sum up, I have greatly benefited from your course, directly and indi-ectly. Most of all, it has made me aware of the importance of making udents more confident about their capacities. I am glad I was in Edin-urgh at the time.

ours

LD

MY ACTION PLAN

will definitely ...

. use / introduce the "getting to know-you"-activities with / to my trainees

try to work more "openly" with poems

try to show my trainees possibilities how to develop learner independence

also hope to ...

include drama activities into my EFL lessons

think more about caring for the needs of the "differently gifted"

/ if wanted: pass on the ideas to my colleagues

ME ...ANNETTE HOLTZ.. DATE October 21, 1994

DRESS

GAERTNERSTR. 87
25469 HALSTENBEK

LD

Dear Jan,

Better late than never - for quite a while I had already intended to write to you, but ... I hope you don't mind me writing by hand, but for me that's much quicker than using any machine whatsoever - and more personal, too. I'd really like to apologize for not having reacted earlier and thus causing your "bad" feelings. I can imagine how disappointed you must have felt after you had put such a lot of effort into your lessons. But let me tell you this - they weren't wasted, at least not on me. Nevertheless, first of all, I owe you some explanations. You see, at the moment (and before) I am (was) doing 4 1/2 jobs at the same time: there's school (19 periods per week), the seminar, a night class, some tuition - and work for a publisher's, which means preparing lectures for colleagues who are interested in a new English book for beginners. All of that means a lot of work - and when your first letter arrived I was up to my neck in it. At the end of January the first half of the German

school year was finished. So, that meant "last" tests to be written, corrected and marked - and lots of conferences after school to decide on pupils' reports. At that time a new semester for my 13 trainees started as well, so get again organising dates, school visits, out-of-seminar activities etc. etc. And finally I got the material for the new book - having to really plough through it to find out how best to present it - which is what I'm doing this month, having ~~to~~ travel ~~to~~ 4 different cities in the North of Germany. Besides, I have started a computer course to really get into contact with that ..... machine.
I find that my usual daily routine looks as follows: getting up. WORK, feeling terribly tired, going to bed.
But to be quite honest, I need the money (I don't get paid for my work at the seminar), as this year I have to face a loss of more than Ƶ 1000,- because of insurance (demanded by the government), taxes and the so-called "solidarity-fee" for the new "Bundesländer". So, one has to ~~try~~ and make ends meet.
But enough of my complaining ....
back to your questions on my "action plan" etc.

just to answer your first question -
frankly, I was and still *am* motivated
by the ideas presented to us in your course.
But there's a difference in my attitude to
being a "teacher" compared to the one expressed
by your "first" writer. I have always tried -
and will go on doing so - to be *open* to:
a) my pupils' needs (whatever age-groups)
b) any new ideas as far as teaching EFL
is concerned
c) new methods in teaching.
I'm of the opinion that a so-called "good"
teacher ought not to think of him/herself
as being the centre of wisdom etc., but always
should try to be open to criticism from others
- and even more important, in my eyes -
be self-critical.
So, which points of my action plan have
worked so far?
- I used a lot of "your" warming-up activities
in all my classes, "punching" ideas for
language learning, even. Especially
my trainees were "delighted" as they're
constantly on the look-out for new methods.
- I haven't had much opportunity for
working with poems (the topics I had to
deal with just weren't right), but I
did songs concerning actual topics
- I myself watched my lesson plans for
enabling my pupils to develop more

learner independence (the new book is a good example has to do that as far as the pupils themselves are concerned)

- and I find myself more prepared to care for the needs of the "differently gifted".

All in all, I can say that, especially after Christmas, I've spent more time to prepare tasks for the pupils, which don't involve me as a teacher, but as a guide and general helper.

Well, to come back to your first writer's statement - I didn't want or intend to become an entirely new teacher, but I <u>do</u> hope that I will constantly be in a process of being able <u>to change</u>, for the sake of my pupils and the benefit of my own personality. (Otherwise, I fear, there's no chance of escaping the "burn-out syndrome".)

In order to make up (a little) for your disappointment, I include some texts I hope you'll not only like but perhaps even will think suitable for one of your courses.

- one page from the book "Ad lib" by Lineaber (EFL for night classes)
- an article from "Hamburger Abendblatt" (a well-known quality paper here in the North), talking about PC
- + my translation (which I hope I did as adequately as possible).

That's all for now. I hope you're a little comforted — and "good luck" for your thesis. Kindest regards to you and your wife. Yours, Annette.

LA

MY ACTION PLAN

I will definitely ...

1. not any longer accept every task / job I'm expected to do at my school apart from teaching
2. concentrate more on making my students talk, play, take part in communication activities
3. inform myself about and try out new approaches like body sculpture or Cuisenaire rods

I also hope to ...

4. be able to interest other teachers in what I had the opportunity to learn
5. keep in touch with the group to perhaps receive further information
6.

NAME BARBARA FISCHER DATE October 21, 1994

ADDRESS

Kappara, April 22, 1995

LA

Dear Jan,
what actually happened when I got back home was that I was confronted with the well-known daily routine again.
(You see that the last question of your second letter is easiest for me to answer – even though I'm on holidays now and should have time enough to reflect more on your points.) But let me just write things down as they come to my mind – without a dictionary, by the way.

I do apologize for not having written earlier. This is not on account of lack of interest but due to too much work. As I might have told you I am among the founding team of one of the newly established comprehensive schools of Schleswig Holstein – a new type of school that in the meantime has been lacking sufficient support by the government. This means really hard work and very little free time. But, of course, that is only one reason for my writing so late. The other one is that I have found it hard to reflect on the impact the course in Edinburgh had on my teaching habits although I feel I profited a lot from it. The question, however, is in what respect.

I really am grateful for the action plan you made us fill in. It definitely helped me to be more decisive in telling people what I'm willing and what I'm no longer willing to do. Which led to the fact that two of my colleagues took over jobs I had been doing since the days when we were only seven teachers and that could meanwhile have well been carried out by other people. To me it means more time for teaching.
I also informed my English-teaching colleagues about the session on poetry we had and at least one of them spontaneously said she wanted to use these texts and ideas as well.
My second attempt of passing on what I learnt about 'drama in the classroom' had to be postponed because of the usual time problems.
When I got back home I was feeling rather euphoric, indeed, and the feeling even lasted for a certain time. I tried out 'body-sculpture' in one of my courses which, in fact, turned out to be quite successful although at first the group felt rather embarrassed and didn't quite understand what I was up to.

Another thing that had immediate impact on my teaching habits was the emphasis I started to lay on communication activities in the foreign languages. (I do French/teach French, too).

While thinking all this over I realize that I have again given in to the daily routine, though.

My answer may be too late for your purposes, I'm afraid, but for me it turns out to be just in time.

Evidently it is worth while looking at one's action plans from time to time!

Please, accept my apologies for the delay of my letter. I'm not so sure that it is of any help to you, but be sure that the course in Edinburgh has been of much help to me.

Yours

APPENDIX 8.4

Further extracts from participant responses to follow up

Comments on non-(immediate) response:

3.1 'Now that I know I'm not the only one being delayed in response I don't feel so bad' (i.e. second letter stimulated response) (E)

3.2 'back home I am so absorbed in everyday routine and tasks that there is no time left ... The course hasn't been unsuccessful nor have I been too lazy. Teaching obligations and duties leave little or no time for extras. In my opinion it's a general déplorable state of affairs. But that won't refrain me from attending courses again as they do make an appreciable difference to me before and after the course' (E)

3.3 (problems with elderly parents) 'I try to work hard but I never have the chance to look inside me' (G)

3.4 'please don't take our non-reactions personally. Trying to understand the four reasons [reference to my second letter] you are on the right track, but there is more to it ... 'too busy enjoying ourselves' - you must be kidding! e.g. I have got to write more than 200 school-reports (i page each!!!) twice a year. And we do individual [?] at school - fun, but an enormous amount of work. Many of us feel terribly exhausted. (J)

3.5 'I have found it hard to reflect on the impact the course in Edinburgh had on my teaching habits although I feel I profited a lot from it' (A)

Further insights into the effect of course process/Ps' feelings during course

3.6 (led to awareness of own lang difficulties) 'I forgot to keep on learning English while I was teaching' and now feels that she is reduced to the productive level of her own students; this affected her participation in the course (G)

Miscellaneous effects of the course:

3.7 (It made me) 'feel more confident that I am on the right way ... to find out that my aims of teaching are right' (G)

3.8 'the mere fact that I attended this course had a general positive effect on me and my way of teaching. It restored much of my initial enthusiasms, which had begun to slumber under years of having to teach English to children who still have difficulty in mastering their mother tongue' (H)

3.9 'All in all, I can say that ... I've spent more time to prepare tasks for the pupils which don't involve me as a teacher but as a guide and general helper' (D)

310 'brought a lot of good ideas back home which I eagerly put into practice without the least delay' (H)

3.11 'Personally, I profited far more from the way and skill you structured and presented your "input" than from the content itself. Surprised ?' (K)

3.12 'I have used some of the material you have given us, but I did not try to imitate you or your approaches' (C)

3.13 'profited more than I thought and in a different way, being more prepared to reflect [on] what I'm doing and how I could change my teaching habits' (I)

In British Council. 1995. *Evaluation in Planning and Managing Language Education Projects*. Second PRODESS Colloquium, 26-28 March 1995. British Council: Manchester

Theme 1 — Evaluation in the Planning Stage of Projects

Some thoughts on project evaluation

Ian McGrath, University of Edinburgh, UK

The gap between theory and practice

I begin with three quotations: the first from ten years ago, the others from PRODESS publications in 1994:

> *Today ... there is an increasing understanding of the importance of objectifiable evaluation of courses and the methodology, content and interpretation of curriculum evaluation is becoming developed and understood. Even in applied linguistics and language teaching, the central importance of evaluation is being recognised and publications in this area are increasing in number.* *(Alderson 1985: 129)*

> *... it would be premature to claim that ... evaluation is an accepted integral part of ELT activity since there is resistance to its use, it is carried out inefficiently, and often is just not used.* *(Murphy 1994: 84)*

> *... it is still fairly uncommon to find project designs that have an in-built evaluation strategy right from their inception.* *(Potter 1994: 7)*

Given the optimism of the first quotation, the comments by Murphy and Potter may seem at first sight rather puzzling. The predicted increase in publications on evaluation has taken place, if at a slightly lower rate than Alderson predicted (see Rea-Dickins 1994 for a survey and Note 2 for a number of significant publications). However, it has not been accompanied by the kinds of change that might have been expected in attitude or approach to project design or course design [1].

This paper offers an analysis of certain facets of the problem and makes a number of proposals in relation to evaluation strategy.

The concept of evaluation for accountability

Most project evaluations in the field of ELT are likely to relate to two or more of the following:

- products;
- teacher development;
- sustainability.

Project objectives refer to them and they subsequently serve as foci for evaluation.

If we adopt a purely quantitative perspective, and quantitative indicators are an element in project frameworks, there should be little difficulty in determining whether a project has met its targets in respect of outputs such as products. Product X - a syllabus or a set of teaching materials, say - is or is not produced by the specified date and within budget. Or, if we take a quantitative approach to teacher development, a specified number of teachers passed through a specified number of training courses. No one would wish to argue that for purposes of accounting, as well as accountability, this information is necessary.

Also necessary, and this applies whether the project has met its targets or not, is what we might call *process* information: records of interim progress, of meetings, of decisions made and the rationale for these, and more subjective data, such as accounts of unexpected difficulties and how these were tackled. Documentation of this kind will be invaluable when project targets have not been achieved,

since it will be important to know why, but evidence of successful practice is equally valuable in informing decision-making concerning the extendibility of projects, and may be helpful to those designing projects in the same region or projects with similar objectives elsewhere - always provided the information is in an accessible form and in the public domain.

We also need a third kind of information, relating to quality. X may have been produced, but did it meet the product specification - was it as good as it was supposed to be? And how did it stand up to subsequent testing? Did it do the job it was intended to? In the way it was intended to? Did it continue to perform satisfactorily over time? (These questions, with a little pronoun modification, can also be applied to teacher development objectives.) If the answer to any of the questions is "No", then again we will want an explanation.

I list below some of the reasons why certain types of information on process or quality may not be collected or, if collected, are not open to inspection, and potential problems with the information that is available.

Information on process

1 No one thinks to collect this kind of data. It is not built into the project specification.

2 It gets lost or is incomplete.

3 It is false: invented to please those above or protect those below.

4 It is edited because too sensitive. Personal, professional or political delicacy acts as a filter on what is officially recorded.

5 It is biased, in that it comes from only one source or a limited number of sources. Certain major stakeholders are not represented.

Information on quality

6 The collection of this kind of information is not specified.

7 The project time-frame does not allow for consideration of effects or sustainability conceived in these terms.

8 No criteria are available for the assessment of quality; no baseline data is collected at the outset of the project.

See also points 3-5.

As I have tried to indicate, evaluation for accountability need not involve a preoccupation with statistical indicators, but it may do so and at the expense of other, possibly more significant data. Another potential disadvantage of an undue regard for final products rather than the processes that lead up to them is that evaluation is not built in from the start of the project; hence, by a series of consequences, the data on which summative judgements are made is limited and the judgements themselves are therefore suspect, since judgements can only be as sound as the data on which they are based. This ought to be a concern, especially when assessments are being made of sustainability or there is a possibility of the project being extended. A recognition at project design stage of the importance of data on process and quality might make a difference, at least in respect of points 1,2,5,6,7,8 above. This should in turn lead to consideration of the skills needed by those who would be involved in the collection of such data.

The finiteness of projects

Projects are, by definition, short-term, and yet they are normally intended to set in motion long-term effects at different levels from the main focus of the project. Such effects are forecast in what may be termed the 'Wider Objectives' cell of project frameworks.

At least two points worry me as far as the evaluation of effects is concerned.

One issue has to do with the level at which one is looking for effects. Chambers and Erith (1990) argue that three levels of desired effect can be distinguished within project frameworks, but that it is only feasible to evaluate effects at the level at which the project has been pitched, i.e. in terms of its 'Immediate Objectives'.

The second point is that it may even be difficult, as noted under Point 7 above, to evaluate immediate objectives when these effects only manifest themselves fully in the long term. This is ironic because without consideration of long-term effects, there can be no meaningful evaluation of the project's overall success.

In this connection, it seems to me important that when outsiders are involved in evaluating a project they should see it as a duty to clarify stakeholders' views on the criteria by which the success of the project can be ascertained and, where these criteria seem vague or unrealistic, their right to re-negotiate them. (This aspect of an external evaluator's role is discussed in Beretta (1990); some of the political factors - and pressures on the independent evaluator - are touched on in Parsons (1990).

Evaluation for development and education

I would not like to give the impression that I am only interested in evaluation for accountability. It is, however, a logical element in project evaluation and, I would suggest, should be a concern - if not necessarily the primary concern - of any professional undertaking. For me, this professionalism extends beyond the taking of quantitative measures to demonstrable attention to processes and quality.

One benefit that accrues from the systematic collection and regular analysis of data on processes and quality is that this information can be used for purposes of *development* as well as accountability. The monitoring activities that tend to happen as a natural part of all projects do, of course, yield feedback that influences subsequent decisions. What is at issue, though, is whether that monitoring is part of a *co-ordinated* evaluation strategy.

My own feelings about evaluation are best captured by the phrase 'evaluation as education'. The term is used by Cumming (1988, after Wise 1980) in a response to a paper by Mackay (1988). Mackay, 'the consummate pragmatist' according to Cumming (1988: 43), presents evaluation as a service to the sponsor; Cumming prefers to see it as a source of learning for anyone who reads the report. The logical extension of this position is that the richer and more insightful the report, the greater is the potential for learning, hence the argument in a previous section for the collection of process data. There are, of course, important issues here: who has access to the report? how many versions of this there are? who owns the data and reports? and so on. These are issues that are best clarified at the outset of a project or when decisions are taken as to who will be responsible for reporting evaluation findings.

Involvement in the evaluation process that eventually leads to a report may also bring positive benefits. As Murphy (1994: 85) notes: 'The lasting benefit from learning to do evaluation is ownership of the means to development, as well as the products of any development'. If this is an intended benefit, then perhaps it needs to be planned in, rather than seen as a spin-off effect.

Moreover, there are what we may call benefits by association. Although there is now more guidance available for would-be evaluators than was the case even five years ago [2], there are few training

courses[3], and therefore the distillation of the experience gained (warts and all) through, for instance, meetings such as the PRODESS Colloquia and the PRODESS Newsletters offers others the possibility of sharing in and learning from that experience.

Evaluation and time-management

Since time is a particular concern within projects, it is disconcerting that it sometimes appears to be spent rather carelessly. I have a friend who says she spends the first 45 minutes of each day at the office planning what she will do and how long she will spend on each item (she sets aside a little time for the unexpected). She claims that this is time well-spent, that she is now able, as a result of experience, to set herself realistic objectives and get through them. This, I suppose, is what is called time management rather than crisis management. I thought of this friend recently as I was reading articles by Ainscough (1994) and Holliday (1992, 1994).

Ainscough provides a frank account of a materials development and INSET project in the Congo in which there was a series of clear accommodations from initial plans to less ambitious but more locally appropriate solutions. One of her conclusions is that more time should be spent on 'understanding the methodologies in use and reaching consensus on aspects of curriculum change' (1994: 19). A second conclusion is in line with the proposals made above for evaluation during the project and for evaluation of longer-term effects. Holliday draws attention to the difference between what he calls 'formal orders', or official reality, and 'informal orders', the reality experienced by practitioners (1992: 406) and points out how difficult it may be to obtain reliable information at the level of informal orders. And yet without such information, as Ainscough recognised by the end of her involvement in the Congo project, one is proceeding on false premises - and wasting time.

The point is that time spent at the beginning of a project (or in a feasibility study prior to a project) is well spent if it:

- reveals the realities of a situation and pinpoints likely difficulties, practical and attitudinal;[4]
- establishes a basis of shared information and understanding between expatriate and local project personnel;
- leads to decisions about not simply what to do but how to do it.[5]

The main problem is perhaps one of perception. At the start of a project, everyone feels under pressure to do something, and with hindsight doing something probably means taking over-hasty decisions. It might be preferable if at an initial stage expatriate project staff spent more time trying to understand, through observation, listening, asking questions, and documenting their understanding (as well as checking that enshrined in consultants' reports). This would be a basis for dialogue. It is also evaluation.

A year or two ago, I was asked to act as a consultant to a project which had not at that time started. In my report I suggested that certain evaluation measures be built in. I was subsequently informed that although the desirability of my proposals was acknowledged, it was no longer feasible to incorporate such measures.

I do not know why, in this particular case, evaluation of the kind that I was proposing had not been built into the project design. I do know that one of the reasons why evaluation does not happen is that there is reluctance on the part of institutions, sponsors and aid donors to accept the related costs. We know from our own experience that when it comes to allocating time we give a higher priority to the development of courses or materials than to their evaluation because that is what our survival instinct dictates. We cannot blame donors if they take the view that money should be allocated to primary, urgent needs. We can, however, try to manage our own time better and persuade donors that money spent on evaluation - if that evaluation is properly prepared for and focused - can actually save time, and possibly money, as well as a good deal of frustration.

Notes

1) See McGrath (forthcoming) for a survey of evaluation practices and attitudes in UK institutions offering INSET programmes in ELT.

2) See e.g. Alderson and Beretta (eds) 1992; Rea-Dickins and Germaine (1992); Weir and Roberts (1994).

3) Mackay (1994).

4) On attitudes to evaluation, see Murkowska in Kiely and Murkowska (1994).

5) See Holliday (1992) on 'means analysis'.

References

Ainscough V (1994) **'Sustainable Methodologies: A Call for More Sensitivity to Relevance in ELT projects.** *In The Teacher Trainer* 8/2: 16-19

Alderson J (1985) 'Is There Life After The Course?' In Alderson J (ed) (1985) **'Evaluation'. Lancaster Practical Papers in English Language Education 6:** Oxford: Pergamon 129-50

Alderson J and Beretta A (eds) (1992) **Evaluating Second Language Education.** Cambridge: Cambridge University Press

Beretta A (1990) **'The Program Evaluator: The ESL Researcher Without Portfolio'.** *In Applied Linguistics* 11/1:1-15

Chambers F and Erith P (1990) **'On Justifying and Evaluating Aid-based ELT'.** *In English Language Teaching Journal* 44/2: 138-143.

Cumming (1988) **'Educating through Evaluation: An Idealist's Response to Ronald Mackay's 'Program Evaluation and Quality Control'.** *In TESL Canada Journal* 5/2: 43-50

Holliday A (1992) **'Tissue Rejection and Informal Orders in ELT Projects: Collecting the Right Information'** *In Applied Linguistics* 13/4: 403-424

Holliday A (1994) **'Overseas Projects: How do we find out what we want to know?'** *IATEFL Management Special Interest Group Newsletter* 16/24-25

Kiely R and Murkowska A (1994) **'Initiating Evaluation Activities: Insider and Outsider Perspectives'** In Kiely, R., Murphy, D. and Rea-Dickins, P. (eds) (1994) 60-72

Kiely R Murphy D and Rea-Dickins (eds) (1994) **'Managing Change and Development in ELT in ECE'.** PRODESS Colloquium, 7-9 November 1993, Manchester. Manchester: The British Council

Mackay R (1988) **'Position Paper: Program Evaluation and Quality Control'** *In TESL Canada Journal* 5/2: 33-41

Mackay R (1994) **'Making evaluation work'** *In PRODESS News 6.* Newsletter of the Project Development and Support Scheme: 5-6

McGrath I (forthcoming) **'Evaluating In-Service Programmes for Language Teachers'** PhD thesis. University of Edinburgh

Murphy D **'Response to Theme 3: Evaluation'** In Kiely R, Murphy D and Rea-Dickins P (eds) (1994) 84-87

Parsons C (1990) **'The politics, ethics and diplomacy of reporting educational evaluation'** *In Evaluation and Research in Education* 4/3: 147-154

Potter M (1994) **'The Way Forward'** In PRODESS News 6, Newsletter of the Project Development and Support Scheme: 7

Rea-Dickins P (1994) **'Evaluation and English Language Teaching'** State of the art article. *In Language Teaching* 27/2: 71-92

Rea-Dickins P and Germaine K (1992) *Evaluation*. Oxford: Oxford University Press

Weir C and Roberts J (1994) **'Evaluation in ELT'** Oxford: Basil Blackwell

Wise R (1980) **'The evaluator as educator'** *In New Directions for Program Evaluation* 5: 11-18