

Durham E-Theses

*The makings of meanings: the evaluation of an
employment programme for drug-using offenders*

Paul Burlison

How to cite:

Burlison, Paul (2003) *The makings of meanings: the evaluation of an employment programme for drug-using offenders*. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a <https://etheses.durham.ac.uk/id/eprint/4051/> is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

THE MAKINGS OF MEANINGS
The Evaluation of an Employment Programme for
Drug-Using Offenders

Paul Burlison

A copyright of this thesis rests with the author. No quotation from it should be published without his prior written consent and information derived from it should be acknowledged.

**A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy**

**Centre for Applied Social Studies
Department of Sociology and Social Policy**

University of Durham

2003



19 JAN 2004

University of Durham

ABSTRACT

THE MAKINGS OF MEANINGS

by Paul Burlison

This thesis is based on a case study of the evaluation of an employment programme for drug-using offenders. Initially, my involvement was that of a studentship-funded research assistant, undertaking the associated duties and expectations of that role and working to an already agreed evaluation design. This design can be broadly categorised as quantitative, longitudinal, and summative. Subsequently, my role evolved, as did the evaluation methodology itself, becoming underpinned by Patton's notion of 'Utilisation Focus', and infused by the, what was initially unarticulated, constructivist hesitation I brought to the endeavour. The findings from the evaluation, which relate to direct Project outputs, staff and service user satisfaction, and service user motivation to participate in the evaluation, are presented. 'Primary intended user' response to these, as well as the evaluation itself, are gauged from interviews undertaken after a reading of the evaluation final report. I then consider what this case study of one evaluation might tell us about the art of evaluation itself. What might be the lessons for evaluation which could be useful for others?



TABLE OF CONTENTS

Chapter 1 – INTRODUCTION	7
Chapter 2 – EVALUATION THE REVIEWED LANDSCAPE	10
Chapter 3 – INTEGRA – THE PROGRAMME	72
Chapter 4 - THE TRUTH I AM TRYING TO GRASP	93
Chapter 5 – NAVIGATING THE ETHICAL SPACE	125
Chapter 6 – ILLUMINATING THE FINDINGS	152
Chapter 7 – PRIMARY INTENDED USER RESPONSE	217
Chapter 8 – DISCUSSION	241
APPENDIX A – CONSENT FORMS	264
APPENDIX B – BALANCE SHEET	269
APPENDIX C – CLIENT INTERVIEW	271
APPENDIX D – STAFF INTERVIEW	274
APPENDIX E – CASE MANAGER QUESTIONNAIRE	277
APPENDIX F – BASW RESEARCH CODE OF ETHICS	283
APPENDIX G – UKES EVALUATION GUIDELINES	287
APPENDIX H – ALCOHOL AND DRUG COSTS	293
APPENDIX I – EXPERIMENT, QUASI-, OR PRE-?	296
REFERENCES	300

LIST OF FIGURES

Figure 2-1: The Realist Evaluation Cycle	65
Figure 3-1: INTEGRA Programme Logic Model I	92
Figure 4-1: Amoeba as Metaphor	97
Figure 4-2: Forms and Functions	107
Figure 6-1: Reasons for Participation	156
Figure 6-2: Comparisons between those interviewed and those not.	159
Figure 6-3: Comparisons between INTEGRA clients and controls on LSI-r	160
Figure 6-4: Social problems: Comparisons between control and INTEGRA samples	162
Figure 6-5: Problem drug use	164
Figure 6-6: Frequency of heroin use by means of administration	165
Figure 6-7: Change in expenditure T2-T1	166
Figure 6-8: Conflict experienced in past month	168
Figure 6-9: MAP symptom inventory I	169
Figure 6-10: MAP symptom inventory II	169
Figure 6-11: MAP symptom inventory III	170
Figure 6-12: Case Managers Survey	210
Figure 6-13: INTEGRA Programme Logic Model II	215

LIST OF TABLES

Table 2-1: Answers to key questions from the paradigm poles	23
Table 5-1: Distinctions between ethical rules, codes, principles and theories	126
Table 5-2: Linking the AEA guiding principles to Kitchener's ethical principles	134
Table 5-3: Linking Kitchener's ethical principles to ethical theories	135
Table 6-1: Number of respondents at T1	153
Table 6-2: Sample characteristics at T1	154
Table 6-3: Number of respondents at T2	154
Table 6-4: Differences between respondents and non-respondents	158
Table 6-5: Differences in LSI-r subscale (item mean) between control group and INTEGRA clients	161
Table 6-6: Frequency of education and employment-related difficulties in control group and INTEGRA clients interviewed	162
Table 6-7: Principal offence category	163
Table 6-8: Total months in prison	164
Table 6-9: Frequency and route heroin and amphetamine use	165
Table 6-10: Expenditure on drugs and alcohol	166
Table 6-11: Balance Sheet analysis from T1.	172
Table 6-12: Work-related outputs	179
Table 6-13: INTEGRA contact	184
Table 6-14: All services used, mean minutes per week	185
Table 6-15: Working status by cohort and interview	187
Table 6-16: Reasons for Closure as recorded in Project database	196
Table 6-17: INTEGRA referral destinations as recorded in Project database	204
Table 6-18: Summary of issues giving rise to positive and negative comments	208

ACKNOWLEDGEMENTS

I am grateful to the many INTEGRA staff and clients who made the evaluation on which this thesis is based possible.

Thanks, too, go to Dr. Justine Schneider, the evaluation manager, and especially to Professors Sue Scott and John Carpenter for suffering the rigours of being my PhD supervisors with such good grace. Sue, you kept me going at midpoint. John, you got me to the finishing line. I will always be grateful.

Wendy, really do miss sharing time and space with you.

Claire and Neil, thanks for being there.

Mother, you've been great.

Josh and Jake, thanks for understanding boys!

STATEMENT OF ORIGINALITY

The data for this thesis were collected whilst I worked as a research assistant on the external evaluation of the INTEGRA Project. The evaluation manager, Dr Justine Schneider, was also my first PhD supervisor, but that role was assumed by Professors Sue Scott and John Carpenter at midpoint.

The core design of the INTEGRA evaluation was established by Dr Schneider before my appointment.

My involvement in the analysis of core quantitative data was through checking data entry on FORMIC, editing, and the production of descriptive statistics. More detailed inferential statistical analyses were undertaken by the evaluation manager, Dr. Justine Schneider, and by Dr. Pauline Coolen-Schrijner from the Statistics Consultancy Unit at the University of Durham. Analysis of all qualitative data, along with the 'Motivation to Participate' and 'Case Manager Survey' components was undertaken by myself.

The presentation and interpretation of the data is my own and does not necessarily accord with the interpretation of the evaluation manager. That said, the basic interpretation and presentation does not differ markedly from that which was negotiated between the manager and myself for the Final Evaluation Report (Burlison and Schneider, 2002).

Chapter 1

INTRODUCTION

*You have to know **why** you want to do evaluation before you can know **how** to do it.*

Saville Kushner, *Personalizing Evaluation*, 2000

I came to this endeavour as a qualified social worker and experienced project manager who had been 'around' evaluation at various points in my professional life. This was both as a designer of small scale internal evaluations associated with posts I have held, and as a stakeholder in external evaluations of programmes with which I have been associated. I have always held the view that a commitment to good evaluation is a commitment to responsible caring. I came to this endeavour to learn more.

Based on the premise that offenders who find employment are less likely to reoffend than those who remain unemployed, the INTEGRA Project aimed to provide a comprehensive service to offenders in County Durham Probation Service (CDPS) whose access to employment was additionally hindered by their alcohol and/or other drug use. Its aim was to increase the number of offenders who entered education, training or employment. The Project was funded by the European Social Fund from June 1998 until June 2000, and linked projects were undertaken by partners in Portugal and Italy. Although working in partnership with local voluntary agencies, the INTEGRA Project was a Criminal Justice project which supervised offenders as a result of community sentences or following release from prison. It was funded by the European Social Fund under the INTEGRA strand of the EMPLOYMENT initiative.

External evaluation of the INTEGRA Project was contracted to the Centre for Applied Social Studies at the University of Durham. This thesis is a case study of this external evaluation. It explores the design, methodology, findings, and dissemination strategy of the evaluation, along with 'primary intended user' response to both the

findings and the evaluation itself. In addition, this thesis also seeks to discuss a range of methodological issues which relate to evaluation as a general process. Integration of such general methodological argument with specific INTEGRA case study reporting is an aim of this thesis, and by so doing I also aim to identify lessons which can be drawn from the evaluation for both.

The thesis structure proceeds thus:

Evaluation today covets flexibility rather than strict adherence to any particular approach. However, this tells us nothing of evaluation's hectic and chaotic past 40 years, or the diversity and passion in the ongoing debate(s) about what evaluation actually is and does. And that is what I am attempting with **Chapter 2**, consideration of the various philosophical and methodological stances as depicted by selected key figures. In addition, I situate my own stance within all of this, preparing the ground for a consideration of my 'approach' to the INTEGRA evaluation in subsequent Chapters. Finally, I offer a personal definition of what I believe evaluation to be.

Chapter 3 reviews research and policy in relation to offender employment issues, and the 'link' between drugs and crime. It explores the Probation Service response to these, as well as European Union employment strategy and responses to the notion of social exclusion. The chapter then describes the INTEGRA Project in detail, illustrating Programme Theory, and introducing themes which underpin key aspects of the evaluation design. The chapter concludes by constructing an initial hybrid programme logic model for the INTEGRA Project.

Chapter 4 is the methodology chapter. It delineates the design and then the process of this programme evaluation, delineates the planning and then the action buttressing its construction.

Chapter 5 discusses associated ethical issues, drawing on the 'Guiding Principles for Evaluators' of the American Evaluation Association, and employing Newman and Brown's (1996) 'cognitive process' framework for ethical decision-making. In so doing, I elaborate on a number of specific ethical dilemmas: 'voluntary informed consent', 'privacy', and 'consent to disclose'; 'disclosure after death'; 'coercion as incentive'; and 'payments for participation'. Of course, this chapter could be synthesised with the previous methodology chapter. However, I consider the themes outlined to be so significant as to warrant their own distinct consideration. This

reflects my belief that ethical thinking is too often limited to, and too often suppressed by, code-driven design concerns.

In **Chapter 6**, I present the major findings of the evaluation, and describe the dissemination strategy for these. The findings address questions relating to Project outputs, staff and service user satisfaction, as well as service users' motivation to be interviewed for the purposes of the evaluation. The Chapter concludes with a revised programme logic model for the INTEGRA Project, based on the findings presented.

Chapter 7 begins by outlining the methodology employed to gauge primary intended user response to the evaluation findings in Chapter 6, as well as to the evaluation itself and the role of the evaluator, and to the evaluation report as a chosen form of representation. I then outline the analysis and key themes emerging from the interviews undertaken, and explore what meanings the respondents created from the evaluation and its report.

In **Chapter 8**, I consider the ways in which philosophy, politics, preference, design, methodology, ethics, and representational forms combine to create the material from which we construct meanings. I also consider what this case study of one evaluation might tell us about the art of evaluation itself. What are the lessons for evaluation which might be useful for others?

Chapter 2

EVALUATION: THE REVIEWED LANDSCAPE

I took a speed-reading course where you run your finger down the middle of the page and was able to read War and Peace in twenty minutes. It's about Russia.

Woody Allen, in Phyllis Mindell, letter to *New York Times*, 3 September 1995

INTRODUCTION

Whilst planned *evaluation* has been noted as early as 2200 B.C., with personnel selection in China (Bowman, 1989), and has been chronicled even more during the last 200 years (see Cronbach *et al*, 1980; Madaus, Stufflebeam and Scriven, 1983), my concern with this Chapter is the field of evaluation today, and how it has developed since the 1960s. For many writers, this period represents the history of evaluation as a 'profession' (see Shadish, Cook, and Leviton, 1991).

In very general terms, evaluation in the 1960s concentrated on experimental and quasi-experimental designs (Campbell and Stanley, 1963). The 1970s witnessed growing attempts to develop evaluations which could be more useful in decision-making. Such attempts have been characterised by Weiss (1987a) as a shift from a knowledge-driven to a use-led approach, or, as Patton (1978) influentially termed it, 'utilization-focused'. The emergence of so-called 'fourth generation evaluation' (Guba and Lincoln, 1989) and the allied 'positivist vs constructivist paradigm wars' of the 1980s bring us, apparently, to a current 'paradigm of choices', a pluralist pragmatic synthesis. Evaluation today, it seems, covets flexibility and methodological appropriateness, rather than slavish adherence to any disciplinary orthodoxy (see Rossi, Freeman, and Lipsey, 1999).

Smooth and unproblematic development then? Just like Woody Allen's sportive speed-reading excursion through *War and Peace*, the previous paragraph, in reality, tells us nothing of evaluation's frenetic and chaotic past 40 years, or the diversity and passion in the ongoing debate(s) about what evaluation actually is and does. But that is what I am attempting with this Chapter, a journey through the various philosophical and methodological meanderings in this chaos, passing en route various scenes of the evaluation landscape as depicted by selected key figures. In addition, I will situate my own stance within all of this, sowing the seeds for a consideration of my 'approach' to the INTEGRA evaluation in later Chapters. Finally, I will offer a personal definition of what I now believe evaluation to be.

At the outset, I ought to say that writing this Chapter has been an unexpectedly perplexing experience. Yes, there is a substantial literature to take on board, and that is quite time consuming. Yes, Woody, it would be wonderful to speed-read. But reading is also the easy bit. Reading the stuff is fine. Thinking about it though is something else. I have been 'around' evaluation at various points in my professional life, either as a designer of small scale internal evaluations associated with posts I have held, or as a stakeholder in external evaluations of programmes with which I have been involved. As such, I thought I had some understanding of, some grip on, this thing called evaluation. But I had not thought sufficiently – either about the purpose(s) of evaluation or the purpose(s) of me. However, the more I think, the less of a grip I realise I have.

So what is evaluation? This question is the starting point of this Chapter's journey. My initial exploration, in *What Is Evaluation? Part 1*, encapsulates the partial consideration that carried me through my participation in the INTEGRA evaluation. But the same question is also posed at the end of this Chapter. My answer there, in *What Is Evaluation? Part 2*, is somewhat different. It represents, what I feel to be, an enriched understanding informed by the additional thinking I needed to tackle this Chapter, additional thinking that occurred once the INTEGRA evaluation was generally 'over'.

WHAT IS EVALUATION? PART 1

Check out any dictionary and the definition of 'evaluation' will invariably involve the notion of assessing the value or worth of something. Why else would the word 'value' be embraced within the body? The Oxford English Dictionary, for instance:

The action of appraising or valuing (goods, etc.); a calculation or statement of value. (OED on CD ROM, version 1.13, Oxford University Press, 1994).

More specifically, given the context of this thesis within a UK Criminal Justice perspective, and for the purposes of the 'What Works' effective practice initiative, the current definition is:

...finding out whether the programme is achieving its objectives (Chapman and Hough, 1998, p.9).

In January 2001, the Home Office, along with the Probation Inspectorate and the Association of Chief Officers of Probation, published a *Handbook for Evaluating Probation Work with Offenders* (Merrington and Hine, 2001) which developed the definition:

...the systematic collection of information about the activities, characteristics, and outcomes of programmes to make judgements about the programme, improve programme effectiveness, and/or inform decisions about future programming. (Patton, 1997, p.23, cited in Merrington and Hine, 2001, p. 2.1).

This definition is a much more useful starting point:

- It emphasises the need for the *systematic collection of information*. This does not mean that a quantitative approach is necessarily required, but that the gathering of information for evaluation must be *purposeful*.
- It specifies a *range of areas* that can be embraced, including process, context, outcomes, and policy.
- It stresses that evaluation *has a purpose*, and is *part of a process*. Evaluation is not an end in itself, but is a means of informing service/policy development and decision-making, and which in turn implies that it must be *used*.

PROVING OR IMPROVING? WHAT'S THE PURPOSE? WHAT'S THE USE?

In outlining their overarching stance to the *purpose* of evaluation, Merrington and Hine (2001) quote Robson's maxim:

...the purpose of evaluation is not to prove but to improve (Robson, 1993, p.180)

This is very laudable. I have to say, though, it came as something of a surprise to see this within the body of a Home Office sanctioned report. Personally, I share Stern's rather pessimistic perspective on UK public sector evaluation purpose:

Evaluation in Britain is shaped by a strong accountability ethic...public sector managers generally use evaluation for two overt purposes: first for management...and second for accountability i.e. to monitor and demonstrate success... the need to demonstrate success tends to dominate the dynamic between evaluators and their sponsors...methods are favoured that try to prove rather than improve... terms of reference for evaluations tend not to challenge policy ...to that extent it is perhaps inevitable that Parliament should value performance above learning. Indeed any administration that admitted learning from its mistakes too readily would be accused of that cardinal sin, a 'U-turn'. (Stern, 1999).

This is a weighty political backdrop for any UK public sector programme, but particularly for one operating within any aspect of the deeply politicised Criminal Justice System.

Evaluation, Evidence, and the UK Criminal Justice System

Andrews (1995) argues that Criminal Justice practice should be based on empiricism. However, in contrast, much Criminal Justice policy has relied more on political and ideological imperatives (Nutley and Davies, 2000). The Conservative Party has long claimed the badge as the party of law and order. Since the late 1980s though, the Labour Party too has laid a vigorous claim to that badge:

The hardening of government policies on law and order during the 1990s can be seen partly as the result of both parties trying to out tough each other in the law and order debate. (Nutley and Davies, 2000, p. 105).

This situation has been compounded in many ways by the lack of a consistent research/evaluation culture within Criminal Justice agencies generally, and within Probation Services in particular. The

House of Commons Select Committee on Home Affairs (1998) concludes:

The absence of rigorous assessment of the effectiveness of community sentences is astonishing... when viewed in the context of the overall expenditure on the criminal justice system, and the further costs of crime both to the victims and to society, the figures spent nationally on research are risibly minuscule. (p. 2).

Moreover, according to McIvor (1995), where evaluation has been conducted it has had limited impact on Probation policies and practices.

In 1999, the Labour Government launched the Crime Reduction Programme, described as 'the biggest single investment in an evidence-based approach to crime reduction which has ever taken place in any country' (Home Office, 1999, p. 3). £250 million was committed to the Programme over the three years 1999-2002.

A number of pre-existing initiatives, aimed at promoting evidence-based policy and practice, have been incorporated into the Crime Reduction Programme, and one of these, as will be outlined in Chapter 3, is the Probation Service 'What Works' project. Pressure is being applied to all Probation Service areas to conduct ongoing evaluations of the effectiveness of their interventions with offenders. Once successful interventions have been identified, once they have 'proven' to be effective, they will form part of a developing menu of effective programmes.

It is not surprising that, faced with such an intense *proving* backdrop, evaluative studies of Probation programmes increasingly tend to adopt a method focus from the positivist traditions. In outlining their understanding of the many current weaknesses in approaches to Criminal Justice evaluation, Nutley and Davies (2000) include:

...a lack of acceptance of the true experiment as an appropriate, practical and ethical means of evaluating many criminal justice interventions... There is a lack of methodological rigour in many existing evaluations of interventions leading to a poor basis for inferring causality (pp. 110-111).

Perhaps the Merrington and Hine (2001) *improving* notion of evaluative purpose marks a sea change in much of this? Then again, perhaps it does not? New Labour asserts a policy commitment to renewal, reform, and modernising Government, with the undergirding 'instrumental' notion of government action informed by reason:

New Labour proclaims the need for evidence-based policy, which we must take to mean that policy initiatives are to be supported by research evidence and that policies introduced on a trial basis are to be evaluated in as rigorous a way as possible. (Plewis, 2000, p.96)

The creation of the Centre for Management and Policy Studies in the Cabinet Office is intended to provide a window at the heart of Government for evaluation and research evidence. However, two recent reports from the Cabinet Office provide clues as to what sort of evidence continues to be viewed as 'gold standard'. A report by the Cabinet Office Performance and Innovation Unit (2000) on improving the role of analysis in policy-making, emphasises the need for 'better' data, 'better' modelling, especially econometric, and more use of longitudinal and experimental designs. Another Cabinet Office report on developing professional policy-making, considers 'new' skills needed by policy makers, including:

....a grounding in economics, statistics and relevant scientific disciplines in order to act as 'intelligent' customers for complex policy evidence. (Cabinet Office, 1999, para. 11.12)

Notwithstanding debate about the role of evidence in policy-making (something which will be considered later in this Chapter when I explore the political and policy framework of evaluation as stamped on the landscape by Carol Weiss, and even if we take-for-granted an 'instrumental' rationalist view of evaluation purpose and use, questions remain about the seeming preoccupation with proving notions of evaluative purpose as driven by an equivalent seeming preoccupation with experimental (and quasi) designs, questions regarding statistical reliability and validity, and a hypothetico-deductive (see later) understanding of causation and generalisation.

'The purpose of an evaluation...

conditions the use that can be expected of it' (Chelimsky, 1997, p.18).

During the first week of November in 1995, in Vancouver, in Canada, 1600 evaluators from 66 countries and 5 continents came together to take stock of where evaluation was at, and where it may be going. 363 panels, workshops, and other sessions, 5 plenary speakers, and a plethora of preconference methodology workshops resulted in a 542 page book (Chelimsky and Shadish (eds), 1997) to 'summarise' the main issues. One thing in particular was clear – evaluation is everywhere, evaluating everything from 'agriculture to zymurgy' (Pawson and Tilley, 1997, p. xi), and even evaluating itself. Evaluation has become a mantra of modernity:

One can begin at the beginning of a dictionary, and go through to the end, and every noun, common or proper, calls to mind a context in which evaluation would be appropriate. (Scriven, 1980, p.4)

Such A-Z diversity is reflected in the differing purposes of evaluation, the differing philosophical bases and methodological emphases these imply, as well as the differing views about the uses of evaluations and evaluators' roles in conducting them. It is also neatly reflected in Wadsworth's (1997) 88-point A-Z catalogue of a bewildering array of evaluation 'models, approaches, and techniques', from 'action evaluation', through 'multi-attribute utility measurement' to 'zero-based budgeting'.

In attempting to grapple with such diversity and summarise the main *purposes* of evaluation, Chelimsky (1997) identifies three general perspectives, albeit not mutually exclusive, and albeit with multiple philosophical and methodological interactions and points of overlap:

- Evaluation for accountability e.g. efficiency
- Evaluation for development e.g. strengthening programmes
- Evaluation for knowledge e.g. better understanding of a topic

To repeat Chelimsky's words at the head of this section, the purpose of an evaluation conditions the *use* that can be expected of it. With regard to the outlined three purpose perspectives, Chelimsky believes that *use* is integrally only a part of developmental evaluation:

...justifying all evaluations by any kind of use may be overly limiting and restrictive for nondevelopmental evaluations (p.18)

This, I have to say, is not my view. Rather, for me, without use there is no evaluation. Without use, evaluation loses the tie to policy that gave rise to the field in the first place. Without use, there may be something else, but it is not evaluation. Use is central to the enterprise, and for that I applaud Merrington and Hine's (2001) definition of evaluation and emphasis on use. I believe that use is also highly relevant as it opens up consideration of 'evaluation' as a distinct activity to 'research'. My own view, again, is that these two terms do indeed denote different landscapes, although I am conscious that this is not a view shared by all. In the Preface to the first edition of her popular book *Everyday Evaluation on the Run* (1991, cited in 2nd edn, 1997), Yoland Wadsworth, for instance, writes:

What this book does as a sequel to *Do It Yourself Social Research* (2nd edn, Wadsworth, 1997) is to examine in far greater detail the *evaluative* elements of research which become uppermost in our minds when we call our research 'evaluation'...to focus separately on research and evaluation is to focus on different elements of an integrated process (p. iv).

The elastic diversity of purpose(s) in evaluation summarised by Chelimsky (1997) may be seen by some as one of its strengths, but it can also be viewed as debilitating baggage, promoting a professional identity, and professional status, for evaluation and evaluators which is unfounded and confusing. Evaluation may have become a mantra of modernity, but at the end of the day, audits are audits, research is research, and evaluation is something else.

Utilisation-Focused INTEGRA

The choice of Michael Quinn Patton's *Utilization-Focused Evaluation* (1997) as the source for the Merrington and Hine (2001) definition, is highly pertinent for this thesis as Patton's framework informed the bulk of my stance in relation to the INTEGRA evaluation (see Chapter 4).

It would be good to be able to say that my choice of Patton was founded on a detailed consideration of the vast literature available to me. But it was not. Yes I scanned other material, but it was pragmatism and serendipity that really drove this choice. Pragmatism in that as an inheritor of an already contracted evaluation, and with the INTEGRA Project already underway, there was little time to navel-gaze with a serious literature search and digestion (at that stage). Serendipity in that Patton's book was the

key text in a 'Policy Related and Evaluation Research Module' convened by our Head of Department, Professor John Carpenter, a module I would undertake at a later stage. Yes, this is pragmatic too, but I prefer the word serendipity as I found this to be a fortunate accident of circumstance – Patton's book is good, his writing style is wonderful, and his framework proved invaluable, particularly as it afforded common ground for myself and my evaluation manager to regard her contracted INTEGRA evaluation methodology as a 'core' which could be added to and developed, with stakeholder agreement (see Chapter 4).

I will outline *Utilization-Focused Evaluation* in more detail later in this Chapter. I will also outline why I now have some problems with it too.

At this point, though, I would like to pick up on the word 'programme' in the Merrington and Hine definition earlier.

Programmes, Programs, and Program Evaluation

My interest with the word 'programme' may seem minor, but it is actually symptomatic of a major issue for anyone undertaking a review of 'evaluation literature'.

In quoting Patton (1997), Merrington and Hine have chosen to edit the word 'program' to 'programme'. This is more than a simple matter of British English chauvinism – although, to be truly chauvinistic, the word 'program' is the earlier 17th Century English and would be strictly preferable to the later 19th Century 'programme' as conforming to the usual English representation in words like *anagram* and *diagram*. Nevertheless, 'program' and 'programme' have become established as the general North American and British spellings respectively¹.

It is no coincidence that evaluation has become commonly referred to as 'program evaluation', where 'program' is a generic term referring to 'some kind of innovation, or intervention, or project, or service' (Robson, 2000, p.8). 'Program' betrays the largely North American origins of evaluation as a discipline, evaluation as an embryonic profession, the North American origins too of many of the programme evaluation pioneers, and the North American context(s)

¹ Some recent and significant British evaluation authors have tended to favour the North American spelling (see for instance Pawson and Tilley, 1997; Kushner, 2000).

in which their ideas and texts have been spawned². It is a North American backdrop with (at least) three caveats to bear in mind when considering the writings of these pioneering evaluators, many of whom went on to develop degree-conferring programmes to transmit their knowledge and develop the profession, and some of whom are still regarded as the 'programme evaluation establishment':

- The early (1960s and 1970s) knowledge base of programme evaluation borrowed heavily from prevailing theories in the social sciences. In philosophical terms, the postulates of logical positivism abound.
- As Shaw (1999, p.4) notes: 'It has been at once the strength and weakness of evaluation that it draws predominantly on the work of education theorists and researchers. The advantages of a shared vocabulary and practical concerns within education have provided the basis for some of the best writing on evaluation... Yet it breeds myopia...'
- Early theorising about the purpose of evaluation, and its situation in relation to the processes of political and social change, has a tendency to focus on relatively large-scale US Federal programmes. Later theorists have developed their magnification to take in local practice. Patton (1997), for instance, mounted his *Utilization-Focused* framework on his work with US community-based initiatives at the local level.

TRADITIONAL ROOTS

Before moving on to chart the 'development' and complexity of programme evaluation, I would like to introduce a brief vocabulary to help communicate some of the philosophical backdrop to what follows.

² Evaluation in the US has developed against the 1960s and 1970s backdrop of major financial investment in social programs initiated under President Kennedy, and expanded under Presidents Johnson and Nixon (Shadish, Cook, and Leviton, 1991, p.22). A major focus was, and remains, the field of education. By the late 1960s, the demand for feedback about social programs exceeded the supply of personnel with appropriate skills. The early demand for personnel had swept up many graduates of professional schools, including accountancy, and social sciences. Between 1960 and 1970 US doctoral production in the social sciences increased 333%, from 2,845 to 9,463. (U.S. Bureau of the Census, 1972, cited in Shadish, Cook, and Leviton, 1991). *Employment in academia did not keep pace with the rapid increase. Professional evaluation became a viable career alternative to the Academy - fountainhead the embryonic profession of 'program evaluation'.*

A PHILOSOPHICAL VOCABULARY

To give adequate thought to the multi-faceted landscape of evaluation, as well as adequate thought to my stance in relation to it all, it is necessary to consider the philosophical underpinnings that give shape to this terrain. Such a frame of reference provides useful markers against which to map key issues and key figures. It also provides a frame of reference to switch on the lights of one's own evaluation understanding and practice. I believe this last point cannot be overstated:

Even more powerful than the notion that there are different paradigms with different assumptions about the world and how it works...is *how much our particular paradigms/assumptions influence the questions we ask; what we think is important to know; the evaluation methods we use; the data we collect; even the interpretations and conclusions we make.* (W.K. Kellogg Foundation, 1998, p. 8)

That said, the question of how far philosophical paradigms impact on evaluation/research practice is contested. This will be covered in more detail later when we consider the notion of 'pragmatism'. For the record, at this point, my position is that paradigms do have significance for evaluative practice. Without some sense of the philosophical underpinnings we bring to any evaluative endeavour, not only do we ignore their possible impact on our contributions to evaluation design and use, but we are also likely to view emergent and unanticipated methodological concerns as mere smudges to be refocused through the lens of our unarticulated instincts.

Intuitively 'positivist'³ evaluators/researchers, for instance, will likely predetermine their problem-solving response by thinking in terms of objectivity, internal validity, and methodological rigour. No matter how 'genuinely pluralistic', no matter how 'genuinely negotiated' with stakeholders, evaluation design is never infallible and it is certainly not coterminous with methodology. Methodological problems will invariably emerge whatever the design.

³ 'Positivist' is a term that now rarely attracts public adherents. Nevertheless, it is a badge claimed by two significant individuals for this thesis: my evaluation manager, Dr Justine Schneider, and Simon Merrington, the co-author of the *Handbook for Evaluating Probation Work with Offenders* (Merrington and Hine, 2001). In an email posting to a 'Probation Practice' discussion forum (21/9/2000) he writes: 'As a positivist (I think) researcher working with quantitative and qualitative methods...'

The first word in this Chapter's philosophical vocabulary, then, is 'paradigm'. Actually, this is quite a problematic word, and more of an awning for a diversity of meaning and description. In very general terms, a paradigm can be said to be:

...a basic set of beliefs that guides action (Guba, 1990, p.17).

However, to expand a little further for my purposes here:

A paradigm is a worldview built on implicit assumptions, accepted definitions, comfortable habits, values defended as truths, and beliefs projected as reality. As such, paradigms are deeply embedded in the socialization of adherents and practitioners: Paradigms tell them what is important, legitimate, and reasonable. Paradigms are also normative, telling the practitioner what to do without the necessity of long existential or epistemological consideration (Patton, 1997, p. 267).

So how many such worldviews are there? Those social scientists who subscribe to a paradigmatic stance (and, as indicated, not all do) are not always in agreement, and do not always employ the same vocabulary. Perhaps there are simply two, the poles of a paradigm world, the 'positivist' and the 'constructivist', undergirding the 'quantitative' and the 'qualitative' respectively. Certainly, in the evaluation world, 'quantitative evaluation' is commonly regarded as depicting the first paradigmatic approach, and 'qualitative evaluation' the second. But is the landscape so simplistic? In his critical analysis of UK contemporary evaluation research perspectives, as specifically applied to the evaluation of practice, Kazi (2000) notes the view of Outhwaite (1987) that there are at least twelve varieties of positivism, for instance, and therefore:

...any single description of positivism as a paradigm will not be able to do full justice to all these variants. (Kazi, 2000, p. 758).

However, such paradigmatic complexity is not necessary for the 'broad-brush' mapping exercise of this Chapter. Rather, I employ the paradigm analysis of Guba (1990) and Lincoln (1990) who distinguish a framework of four worldviews: positivist, postpositivist, constructivist, and critical theory. This is the framework that is

similarly used by Shaw (1999) in establishing the philosophical groundwork for his investigation of *Qualitative Evaluation*.

Consider four philosophical categories and some associated key questions:

- **Ontology** What is the nature of the world? What is reality? What counts as evidence?
- **Epistemology** What is the relationship between the knower and the known. How is knowledge constructed? What role do values play in understanding?
- **Logic** What are the principles of demonstration and verification? Are causal links possible?
- **Teleology** What is the purpose of evaluation? What does it contribute towards knowledge?

Now consider (in Table 2.1 over page) what these key questions can show us about the major postulates of these two paradigm poles.

POSITIVIST OBJECTIVISM AND POSTPOSITIVIST REALISM

For the positivist, 'the business of science is to discover the true nature of reality and how it truly works' (Guba, 1990, p.19). It has an objectivist epistemology, and the idea of an Archimedean point:

Objectivity is the Archimedean point....that permits the enquirer to wrest nature's secrets without altering them in any way. (Guba, 1990, p.19)

Influenced by the philosophical postulates of logical positivism see Table 2.1), the roots of programme evaluation are dominated by natural science notions of, what Patton calls (1997, p. 268) 'hypothetico-deductive methodology'. Stemming from the tradition of experimentation in agriculture, the epitome of 'good science' is regarded as the statistical analysis of the quantitative data arising from experimental designs.

In their widely used methodology primer, Donald T. Campbell and Julian Stanley (1963, p. 3) described this approach as 'the only available route to cumulative progress'.

Table 2.1: Answers to key questions from the paradigm poles

Questions	Postulates of the positivist paradigm	Postulates of the constructivist paradigm
1. What is the nature of the world? (Ontology)	Reality is one. By carefully dividing and studying its parts, the whole can be understood.	There are multiple realities which are socio-psychological constructions forming an interconnected whole.
2. What is the relationship between the knower and the known? (Epistemology)	The knower can stand outside what is to be known. True objectivity is possible.	The knower and the known are interdependent.
3. What role do values play in understanding the world? (Epistemology)	Values can be suspended in order to understand.	Values mediate and shape what is understood.
4. Are causal linkages possible? (Logic)	One event comes before another event and can be said to cause that event.	Events shape one another. Multidirectional relationships can be discovered.
5. What is the possibility of generalisation? (Logic)	Explanations from one time and place can be generalised to other times and places.	Only tentative explanations from one time and place are possible.
6. What does evaluation contribute to knowledge? (Teleology)	Generally, the positivist seeks verification or proof of propositions.	Generally, the constructivist seeks to discover or uncover propositions.
	▼ Quantitative Approaches	▼ Qualitative Approaches

(adapted from Maykut and Morehouse, 1994, p.12 – as they, in turn, adapt from Guba and Lincoln, 1985)

Indeed, Campbell and Stanley's *Experimental and Quasi-Experimental Designs for Research*, and its key concepts of internal and external validity, has been rated in the US as more influential

than any other evaluation work or concept (Shadish and Epstein, 1987)⁴.

Donald T. Campbell – ‘internal validity’ and ‘cause-probing’

While it may seem unfair to underscore one individual from a co-authored work, I make no apologies when that co-author is the polymath Donald T. Campbell. Quoting Chelimsky and Shadish’ (1997) *In Memoriam* preface (Campbell died in May 1996):

Looking back over Campbell’s work, it is clear that he has consistently tried to integrate opposite poles in evaluative thinking, to bridge the gaps between them. This effort of his was crucial to evaluation’s survival over the past decade, when it was under attack, and it promises Campbell an enduring place, not only in the annals of applied social science research, but also in whatever pantheon we have in America for people who can understand both sides of an issue and integrate them into a larger – but always elegant – framework.

Looking back over anyone’s published work, though, is always fraught with difficulty and the potential illusion of tidy continuity and certainty. Meaningful roles are seldom credited to boredom, cul-de-sacs, chance, or lax wandering. Nevertheless, Campbell’s fifty year intellectual career, as described by Shadish, Cook, and Leviton (1991, p.121):

....can be seen as....interest in describing and explaining how humans, including scientists and scholars, learn about the real world and how that learning might be improved.

Campbell’s interest in evaluation was actually secondary to his broader intellectual concerns with psychological theory, methods, sociology of science, and epistemology. His main theoretical contributions are an evolutionary theory of knowledge growth (1960; 1974), and a language for identifying and controlling bias threats to the validity of what he describes as ‘cause-probing’ studies (1957; Campbell and Stanley, 1963). As a result of his early work, he was

⁴ I had hoped, with the advice of Will Shadish and others, to update and internationalise this work for the purposes of this PhD via an online survey of current evaluators’ practice and theoretical influences. However, this was seen by my PhD supervisors as methodologically problematic, time-consuming, and unnecessary for this thesis.

lionised as the father of scientific evaluation. This was more by accident than any design, and owes much to Suchman (1967):

Suchman's 1967 founding book on evaluative research cited my "experimental and quasi-experimental designs" as the appropriate methodological mode, I thus became overnight both a senior programme evaluator by fiat, and one committed to an experimental epistemology and the theory of science shared by the physical sciences. (Campbell, 1984a, p.13)

Suchman believed Campbell's work represented *the* model for evaluation, and it became widely assumed in the 1960s that experimental designs to measure the outcomes of innovative programmes were the benchmark for *all* evaluation.

Since the 1970s, Campbell has been interested in using the history and sociology of science to describe how science can only approximate truth, not simply because of research design, but because of sources of validity and invalidity in the passions and prejudices of individual researchers and research communities. In this, he urged the development of public criticism in evaluation to scrutinise knowledge claims more intensely. He aimed to create a 'disputatious community of scholars' (Campbell, 1984, p.44), one which seeks debate, includes dissenting opinion in reports, funds multiple evaluations of a programme not just one, and regularly resorts to reanalysis of others' evaluation data. His broad utopian vision was of an 'Experimenting Society':

An experimenting society would vigorously try out possible solutions to recurrent problems and would make hard-headed, multidimensional evaluations of outcomes, and when the evaluation of one reform showed it to have been ineffective or harmful, would move on to try other alternatives. (Campbell, 1988, p.291)

In all of this, Campbell is best described as a postpositivist⁵. He accepts the existence of a reality beyond knowers, yet rejects the

⁵ In 'postpositivism' the suffix 'post' conveys the idea that the position keeps some of the key ontological/epistemological premises of positivism, with some departures around the notions of realism and subjectivity. In this sense, 'post' indicates some continuity. This contrasts with 'postmodernism', where 'post' conveys the idea that key motives of modernism are fundamentally challenged, and, as such, it indicates discontinuity.

naive realism in assumptions that our observations can mirror this reality directly. While he clearly prefers quantitative methods, and gives greater weight to internal validity rather than external validity, Campbell also criticises quantitative scientists who:

....under the influence of missionaries from logical positivism, presume that in true science, quantitative knowing replaces qualitative, common-sense knowing. (Campbell, 1979, p.50)

In this way, Campbell recognises the contributions of qualitative methods which he sees as complementing rather than replacing the quantitative. He notes that quantitative measurement rests on qualitative assumptions about which constructs are worth measuring. He also values knowledge obtained directly from programme participants because of their grounding in programme experience. But Campbell also goes further by arguing that qualitative methods can fulfil the same cause-probing function as quantitative methods, particularly experiments. His argument here is complex, and not necessary for my purposes, but suffice to say it distils to the notion that the amount of uncertainty reduced about a knowledge claim depends primarily on the extent to which plausible alternative interpretations are ruled out. The crux is the ruling out of alternatives – how they are ruled out is irrelevant. In 1994 he summarised his appreciation of qualitative approaches thus:

...in addition to the quantitative and quasi-experimental...our social science armamentarium also needs a humanistic, validity-seeking, case study method'. (Campbell, 1994, p.x)

As such, Campbell's work has been used to justify evaluation practice centred on both quantitative *and* qualitative approaches – although it is the former which clearly dominates.

Michael Scriven – 'summative', 'goal-free', and 'value claims'

Like Campbell, the philosopher Michael Scriven has helped shape evaluation theory through four decades. His concern, though, is less with developing evaluation methodology and more with exploring the logical requirements of the task. Like Campbell again, he has introduced a vocabulary that has become institutionalised in the field. Scriven, for instance, is responsible for introducing the summative-formative distinction:

Evaluation may be done to provide feedback to people who are trying to improve something (formative evaluation); or to provide information for decision-makers who are wondering whether to fund, terminate, or purchase something (summative evaluation). (Scriven, 1980, pp.6-7)

Summative evaluations are deductive, whilst formative are generally inductive, with less formal criteria as one searches for whatever strengths or weaknesses may emerge from looking closely at what is happening in a programme.

Scriven, given his mantra that evaluation should assign merit or worth, prefers summative approaches. For him, the idea of evaluators who never make evaluative conclusions:

...is not only paradoxical, but it makes the notion of an autonomous profession of evaluators redundant. They are no more than relabelled social scientists. (Scriven, 1995, pp. 54-55)

He is careful to draw distinctions between judgements, explanations, and recommendations ('remediations'):

Bad is bad and good is good, and it is the job of evaluators to decide which is which. And there are many occasions when they should say which is which, whether or not they have explanations or remediations. (Scriven, 1986, p. 19)

He argues that value claims are similar to other scientific constructs:

Value, worth, quality, and merit are simply constructs from observable variables, just as aptitudes and achievements and motivation and anxiety are. (1986, p.39)

Establishing the validity of value claims is similar to establishing the validity of other scientific constructs, they are not directly observed, but are 'indirectly demonstrated or inferred from the results of tests'. (1986, p.39). The validity of a value claim is a matter of joining factual claims with a network of knowledge claims and searching for consistencies. To appreciate this requires an equivalent appreciation of Scriven's postpositivist perspectivism. He rejects the idea that there is no reality or that it is not possible to describe

reality. But he also rejects positivistic ideas that there are bias-free building blocks of knowledge, where empirical facts are the sole arbiter. No single scientist ever observes reality completely, but rather observes through biased perceptual filters. Using multiple perspectives helps construct a more complete picture of reality. He calls this 'perspectivism':

Perspectivism accommodates the need for multiple accounts of reality as perspectives from which we build up a true picture, not as a set of true pictures of different and inconsistent realities. (Scriven, 1983, p. 239).

Consistent with this perspectivist epistemology, is Scriven's emphasis on a multimodel of evaluation: 'Evaluation is a multiplicity of multiples (1983, p.257):

Evaluation is *multifield*, concerned with programs, products, personnel, plans, and potentials; *multidisciplinary*, with *multidimensionality* of criteria of merit; needing multiple perspectives before synthesis is done; *multilevel*...in the different levels of analysis, evidential support, and documentation appropriate in different circumstances; using *multiple methodologies*, *multiple functions*, *multiple impacts*, *multiple reporting* formats. (Shadish, Cook and Leviton, 1991, p.82).

As a way of instantiating this multimodel, Scriven has developed The Key Evaluation Checklist. Not only is this designed to help evaluators keep on track with his evaluation logic, but also to assist the auditing of the final report – a special case of meta-evaluation, ensuring 'the evaluators get evaluated' (Scriven, 1976, p.126), and ensuring that bias control remains at the centre of the evaluator's agenda to the very end of the enterprise.

Bias control, then, underpins Scriven's entire motif. His ideal means of handling bias, from the very outset to reaching a summative judgement, is his trademark of 'goal-free' evaluation. Goal-free evaluators are 'totally blind' (1976, p.137) to stated goals. Kept in ignorance of a programme's goals, the evaluator will '...struggle hard to find any and all effects, without prejudice, since his or her reputation is on the line' (1976, p.137). Evaluators, then, 'have to discover what effect' the program has, and match 'their effects against the needs of those whom they affect' (Scriven, 1983, p.235).

The goal-free evaluator is totally insulated from any direct contact with project people at all, is not allowed to talk with them about goals or vice versa, is not allowed to get long histories of the project, and so on. He doesn't, in fact, communicate with the people that run it. (Salasin, 1974, p.10)

The distinction between intended and unintended effects is one Scriven rejects as reflecting the intentions of the producer. Striving to be 'goal-free', then, leads to evaluation practice where the lead evaluator is almost inevitably an outsider:

Both distancing and objectivity remain correct and frequently achievable ideals for the external evaluator. (Scriven, 1997, p.483)

'Distance has its price', but involvement 'risks the whole capital', and 'so-called participatory design....is about as sloppy as one can get' (Scriven, 1986, p.488, 486).

As mentioned at the outset to this section, Scriven's concern is less with developing evaluation methodology and more with clarifying the logical requirements of the task in a way that will improve practice. He summarises the logic this way:

....evaluation involves determining criteria of merit (usually from a needs assessment), standards of merit (frequently as a result of looking for appropriate comparisons), and then determining the performance of the evaluand so as to compare it against these standards. (Scriven, 1980, p.18)

Campbell, Scriven and the roots of Programme Evaluation

The roots of programme evaluation are epitomised by Campbell and Scriven. Both are postpositivist realists, believing a reality external to the knower exists. They also believe it is possible to construct more or less valid knowledge about this reality, and they share an interest in practical causal statements about 'what works'. But bias is a constant threat, and so each stresses bias control. Scriven focuses on controls that guard against the evaluator being co-opted; whereas for Campbell the controls lie in knowledge of multiple validity threats, use of experiments, and public debate of knowledge claims. In this, then, they both promote evaluator distance from stakeholders so as to avoid compromising the evaluation's integrity.

For both Campbell and Scriven, a valuable social programme solves important social problems. Part of the evaluator's job, then, is to render some sort of judgement about just how valuable the programme is. Scriven is more explicit and detailed on this point than Campbell (and probably more than any other theorist for that matter). Scriven's crux is to construct value statements, value statements that are matters of fact, not opinion. His logic of evaluation is his method for doing this: evaluators should assess how much the object being evaluated meets important needs, particularly in comparison with alternatives for meeting those needs; then a single value judgement should be rendered about its worth. Campbell follows a similar logic, comparing the effects of programmes to controls or comparison treatments that might accomplish the same ends. Where they part company is at the final step in the logic: Campbell favours allowing stakeholders to render their own judgements of worth.

When it comes to evaluations being used, neither Campbell nor Scriven discuss this extensively. In assessing their attitude to use, Shadish, Cook, and Leviton (1991, p.70) categorise Campbell and Scriven as 'naïve instrumentalists', and infer two characteristic assumptions:

- Feedback about programme effects will be used to maintain or expand effective programmes, and make radical changes to those shown to be ineffective.
- Evaluators have to do little for such use to occur.

Scriven handles use the way market consumer reports do: stakeholders will use information as they see fit, depending on how well it meets their needs. Campbell assumes data is used to improve programmes, and fears that active evaluator attempts to promote use might compromise the standing of findings. That said, he also advises evaluators to monitor how results are used, and to protect the truth even if only in scholarly debates with colleagues.

To summarise, then, Campbell and Scriven may differ in some of their methodologies, with Campbell (generally) being more traditionally scientific, but their work has the same summative output-focused goals, emphasises evaluator distance from stakeholders, has a similar logic, and makes comparable assumptions about use. In all of these, they set the conventional tone for programme evaluation.

ALTERNATIVE PARADIGM SHOOTS

'Evaluation research has not exactly lived up to its promise' (Pawson and Tilley, 1997, p.2). The theses produced by such conventional approaches to programme evaluation were felt to be lacking in their approach to the creation of useful knowledge, in their appreciation of the nature of social change and the policy process, and in the *laissez-faire* attitude of evaluators in encouraging the use of their work. The 1970s witnessed the dawn of serious attempts to find alternatives.

One paradigmatic alternative, of course, already existed...

...another way of studying program processes and outcomes that began to attract a following from evaluators and practitioners who found that the dominant paradigm failed to answer – or even ask – their questions. (Patton, 1997, p.271)

NATURALISTIC, INTERPRETIVE, CONSTRUCTIVIST

This alternative (see Table 2.1 for a summary of major postulates) arises from over a century long critique of the positivist scientific method as a basis for understanding human activity. The nineteenth century German historian and philosopher Wilhelm Dilthey argued that a clear distinction should be drawn between the disciplines of *Naturwissenschaften* (natural science) and *Geisteswissenschaften* (the moral or human sciences). In Dilthey's view ([1894] 1977), whilst the former could be prosecuted by the external observation and explanation of regularities in physical events, the human sciences should be premised upon the search for *Verstehen* (meaning or understanding). The paradigm, as typically characterised today, draws upon developments from these early insights, and their echoes in the traditions of hermeneutics and phenomenology.

Whilst a challenging variety of constructivist positions exist, qualitative methods are generally privileged within all aspects of the paradigm, as is a tendency towards formative approaches. **Robert Stake's responsive approach** is one such early alternative to traditional evaluative endeavours:

...it is an approach that trades off some measurement precision in order to increase the usefulness of the findings to persons in and around the program. (Stake, 1975, p.14)

Although qualitative methods were present in some predecessors of programme evaluation, such as action research and ecological psychology (Marrow, 1969; Barker, 1968), Stake helped legitimise them in evaluation, and catalysed 'debates about epistemology, ontology, and disciplinary myopias' (Shadish, Cook and Leviton, 1991, p.271).

Stake emphasises the use of case studies:

The case need not be a person or enterprise. It can be whatever 'bounded system'...is of interest. An institution, a program, a responsibility, a collection, or a population can be the case. (Stake, 1978, p.7)

Clearly case studies can be statistical, but for Stake:

...most case studies feature: descriptions that are complex, holistic, and involving a myriad of not highly isolated variables; data that are likely to be gathered at least partly by personalistic observation; and a writing style that is informal, perhaps narrative, possibly with verbatim quotation, illustration, and even allusion and metaphor. Comparisons are implicit rather than explicit. Themes and hypotheses may be important, but they remain subordinate to the understanding of the case. (1978, p.7)

In all of this, the evaluative objective of improving *local* knowledge is central to the enterprise. Whilst formal theory and codified data can also affect local practice:

We speak not against such knowings but claim they are too often exclusively relied upon, too regularly presumed to be the preferable messages. If we look on the sketch the leverage point for change too often neglected is the disciplined collection of experiential knowledge. (Stake and Trumbull, 1982, pp.8-9)

For Stake, then, case studies help rectify such historical neglect of vicarious experience in producing change. When it comes to understanding the process of practice development, Stake appreciates the importance of taking on board the developments suggested by others, as well as the role of a better understanding of ourselves and our relationships to others. But Stake also argues that both approaches rely on a third method of change:

One may change by adding to one's experience and re-examining problems and solutions *intuitively*...We maintain that this third method of planned change...is at least as important as the other two and that program improvement efforts should more often rely upon the experiences and intuitions of the practitioners involved. We believe that program evaluation studies should be planned and carried out in such a way as to provide a maximum of vicarious experience to the readers who may then intuitively combine this with their previous experiences. The role of the program evaluator or educational researcher would then be to assist practitioners in reaching new understandings, new *naturalistic generalisations*. (1982, p.2)

In this way, responsive evaluation can help create tacit knowledge for people beyond local involvement with the programme, through the reading of case study reports and then generalising to their own situations to assist improvement.

Cross-fertilisation of evaluative thinking between the US and Britain has been a strong feature of Stake's case study work, in particular the rise of *democratic evaluation* associated with the Centre for Applied Research in Education (CARE) at the University of East Anglia (UEA). McDonald, Parlett, Hamilton, Stenhouse, Walker, Smith, Adelman, and Kushner are among the names associated with CARE, as are Stronach and MacLure, both significant contributors to the development of postmodernist evaluative thinking, something to which I will return a little later in this Chapter.

In addition to Stake, two other names in the above UEA list are also credited with producing a founding text of the alternative paradigm in evaluation. Parlett and Hamilton's (1977) *illuminative evaluation*:

...takes account of the wider contexts in which educational programs function. Its primary concern is with description and interpretation rather than measurement and prediction...it aims to discover and document what it is like to be participating in the scheme, whether as teacher or pupil...to discern and discuss the...most significant features, recurring concomitants, and critical processes. In short, it seeks to address and illuminate a complex array of questions. (p.144)

Although they would probably prefer the suffix 'research' or 'inquiry' rather than 'evaluation' to describe their commitments, we can also add to this paradigm naissance inventory the names of Glaser and Strauss (1967), and Reason and Rowan (1981).

However, no two names in evaluation epitomise more the paradigmatic pole of radical relativist constructivism than Yvonna Lincoln and Egon Guba. Whist Stake may have introduced an interpretive thrust into evaluation theory:

I have increasingly replaced realist presumption with constructivist hesitation (Stake, 1991, p.81),

it is Guba and Lincoln who have changed the way many evaluators fundamentally think. In their seminal *Fourth Generation Evaluation* (1989), they outline their relativist, constructivist, and value-infused paradigm. For them, there is no objective reality, no single 'out there' out there. Rather, 'reality' is entirely a human construction that depends, again entirely, on agreement among participants:

Now constructions are, quite literally, created realities. They do not exist outside the persons who create and hold them; they are not part of some 'objective' world that exists apart from their constructors. (Guba and Lincoln, 1989, p. 143)

Realities, then, cannot be studied as separate bits, as separate variables, but only holistically and in context. For Guba and Lincoln, there is little to choose between the perceptions of different individuals, including the evaluators', and the only hope we have of fixing reference points is by negotiated agreement:

...the relationship, when properly established, is one of respectful negotiation, joint control, and reciprocal learning. (Lincoln and Guba, 1986, p.75)

In this way, the knower and the known are 'fused into a coherent whole' (Guba, 1990, p.26). And because all behaviour is context and time bound, Lincoln states in her personal account of her journey to constructivism:

...we began to doubt seriously the possibility of generalization from one site to the next. (Lincoln, 1990, p.68)

Their model for evaluation is the 'hermeneutic dialectic circle' of inquiry. Major stakeholders are first identified and their concerns and beliefs solicited. Second, all stakeholder views are submitted to other stakeholders for comments and criticism. Next, those issues that are not resolved by this initial discussion become the organisers for data collected by evaluators in the next stage. Finally, this collected information is considered by all stakeholders in joint discussion, and there is an attempt to reach consensus on each disputed item:

Views are neither right nor wrong but different depending on how the construction is formed, the context in which it is formed, and the values that undergird construction in the first place. It is not more research that is needed but more negotiation. (Guba and Lincoln, 1989, p.255)

As such, Guba and Lincoln do not claim that positivism is wrong or untrue, rather it is 'ill-informed and unsophisticated':

The relativist constructivist, while not agreeing with the positivist formulation, can nonetheless accept it as one of many constructions. The constructivist may find the positivist view ill-informed and unsophisticated, but not wrong or untrue. (p.16)

The negotiation process ends when consensus is reached, or time and resources run out. In this way, evaluation creates the reality that it presents, and the evaluation report is simply the residue from this. Some issues, of course, will likely not be resolved. But for those that are, action is implied:

By their joint agreement they commit themselves jointly to accept continuing responsibility and accountability for whatever action is taken. (p.222)

During this process, 'Evaluators are orchestrators of a negotiation process that aims to culminate in consensus on better informed and more sophisticated constructions' (p. 110).

Evaluators insert their own opinions for consideration only later in the dialectic process, save too much weight be given to them, and then, too, in a way which does not identify the source:

...we think that the evaluator's construction would be invested with too much weight if announced to be that. However, if the evaluator's construction is introduced...in some low key way – 'Some people believe that...' – it is open to critique without fear of reprisal or embarrassment. (p. 213)

In Guba and Lincoln's view, the chance that evaluator biases will shape the conclusions is low, provided that the process is conducted in accordance with hermeneutic dialectical principles.

I should say at this point that there is much in the work of Guba and Lincoln, as well as Stake, that I find personally appealing. After much philosophical navel-gazing, I generally feel that I share their constructivist perspective. In this way, for instance, I comprehend evaluation itself to be a personal construct, something to which I will return at the end of this Chapter. However, unlike Guba and Lincoln, I am (usually) unable to grasp the relativist nettle so tightly, I do not see relativism and constructivism as twin sisters. Indeed, I do not see constructivism as being wholly incompatible with aspects of realism. Whilst I do not recognise objective reality out there, I do recognise the existence of embedded social systems, and, with evaluation firmly in mind, this is no more apparent than when considering social policy and social programmes. I also share a realist's sense that fallible knowledge of these social systems is attainable. My constructivism, though, centres around the contextual construction of this knowledge, the importance of values and subjectivity, my unease around causation, and my uncertainty regarding transferability to other settings. Without constructivist hesitation, I believe that we do not necessarily and progressively participate in some sort of critical realist 'natural selection of knowledge', but that rather we run the risk of colluding in some sort of 'natural selection of attractive ignorance'.

My main problem with Guba and Lincoln lies in, what House and Howe (1999, p.69) describe as, their 'hyper-egalitarianism'. If, as Lincoln and Guba (1986, p.79) state:

...all ideologies should have an equal chance of expression in the process of negotiating recommendations,

and if, as they state, 'it is the mind that is to be transformed, not the real world' (Guba, 1990, p.27), then the potential for critical or reformist action seems, at best, contingent:

The essential relativism of interpretivism argues for no particular role in the world. There is no commitment to reinforce the status quo or to challenge it, to condone racist or sexist practices by one's silences or to actively seek to redress observed inequalities, to remain neutral on issues of political beliefs or to openly advocate for a chosen ideology. (Greene, 1992, pp.43-44)

In this way, relativism may have taken a lion's bite out of other philosophies, but it is a mouse when it comes to political action. This is an issue which raises fundamental questions regarding the role of the evaluator and the purpose of evaluation. For Guba and Lincoln, the evaluator is an orchestrator whose constructions are given no more weight than other stakeholders. Whilst I applaud the participatory and emancipatory ideals underpinning this view, and whilst I also recognise the growing emphasis of such evaluative approaches in developing countries, I also feel it carries a risk of political complacency. The issue here is one of power. If the means for equalising power is simply to check the authority of evaluators, then this minimises the added weight evaluator constructions may well hold as a consequence of their knowledge base and experience. But, and more importantly in my view, it also denies the capacity of evaluators to redress power imbalances which may lie amongst the stakeholders themselves, as well as their capacity to challenge stakeholder constructions which may be unethical or contrary to notions of social justice. This opens consideration of the evaluator role as a reformist advocate:

Truth is more important than beauty. And justice more important than either. (House, 1980, p.117)

The ethical dimensions of this will be discussed in more detail in Chapter 5. For now, I would like to explore this view of reformist advocacy within another paradigm alternative to the conventional view...

CRITICAL EVALUATION

In many ways, I wish I had not followed so faithfully the paradigm taxonomy of Guba (1990) and Lincoln (1990), and included this fourth worldview of 'critical theory', or, more specifically for my purposes, this worldview of 'critical evaluation'. As Shaw (1999) notes:

...there is still no prominent theorist of 'critical evaluation' who will 'stand for' the application of critical theory to evaluation. (p.33)

Without reference to a prominent illustrative theorist this section could become very unwieldy indeed. Of course, the name Ernest House (1980) springs readily to my mind, as indeed it does to Shaw, but we should exercise caution:

According to House's reformist view, disagreements between the stakeholders of evaluation occur within a general consensus regarding fundamental democratic values. He noted emerging critical models of evaluation with an acknowledgement that any such persuasion regarding evaluation would stand outside his own framework for understanding the political location of evaluation. (Shaw, 1999, p.33)

Critical evaluation has essentially become an umbrella, under which shelter various 'openly ideological stances' (Greene, 1997), and which advance various ideals, values, or social change agendas. A, not by any means exhaustive, list of examples would include: neo-Marxist evaluation (Anderson, 1989; Lather, 1991; Popkewitz, 1990); some feminist positions (Humphries, 1999; Swigonski, 1993); participatory evaluation (Whitmore, 1994; Reiben, 1996); and empowerment evaluation (Fetterman, 1994; VanderPlaat, 1995).

These approaches may differ in the nature of, and justification for, the agendas they promote, but they share a privileging of their stance over the demands for technical rationality, and they stress attention to the political and moral dimensions of the evaluator's role and responsibilities. They are 'critical' in the sense that evaluation problems are conceptualised within their social, political, and cultural context. It is 'critical evaluation' because it 'gives reference to a systematic inquiry that focuses upon the contradictions of practice' (Popkewitz, 1990, p.46). The basic logic, then, is not solely preoccupied with the formal organisation of argument, 'but also particular forms of reasoning that give focus to scepticism towards social institutions' (p.49), and sometimes towards empirical evidence, 'thus poking holes in the causality that confronts us in daily life and that limits our possibilities' (p.49).

Critical knowledge from evaluation, then, is never neutral. The production of knowledge is the production of values. This has major implications for evaluators who are:

...participants in the socio-historical development of human action and understanding. As such they must decide which interests they will serve. The only legitimate activity...is to engage in the collective enterprise of progressive enlightenment. (Comstock, 1982, p.377)

There are differences within critical evaluation at this point. Perhaps the majority view is that the partisan role of science makes it an obligation to pursue political commitments through 'a direct and explicit involvement in efforts to transform current social relations' (Popkewitz, 1990, p.49). Thus activism is implicit, and inquiry aims to disclose ways in which people are to challenge the world and locate themselves in its ongoing relations.

Given this sense of 'critical evaluation', then, Ernest House would indeed feel uncomfortable to have his reformist views similarly politically located. However, critical evaluation as framed by Everitt (1996) has a somewhat different flavour. This holds stakeholder agency, equitable power distribution, and democratic dialogue as primary values. These form the standards against which both the social programme being evaluated, and the evaluation itself, are judged. Critical evaluation thereby contributes to 'good' practice, practice that is democratic and fair, and, in part, it does so by democratising the ways in which evaluative evidence is generated.

Accepting Everitt's, albeit less politically strident, frame for critical evaluation, I feel now gives some justification for positing my desired consideration of Ernest House in this section. I will begin this consideration by picking up directly on the notion of advocacy, introduced at the end of the previous section, but clearly implied throughout this section so far.

The unavoidable web of advocacy

Although an issue of contemporary debate, deliberation of advocacy within programme evaluation is not new. Twenty years ago, Ernest House urged that we think of programme evaluation as argument rather than demonstration:

Evaluation persuades rather than convinces, argues rather than demonstrates, is credible rather than certain, is variably accepted rather than compelling. (House, 1980, p.73)

For House, this is as true for quantitative evaluation as it is for qualitative, despite the fact that 'statistical metaphors...give a semblance of certainty and unequivocal to evidence' (p.74).

Stripped of the illusion of positivist certainty, evaluators are inherently caught in a web of potential advocacy. Values engender such advocacy, sometimes subtle, sometimes blatant and deliberate. As our values, both personal and institutional, are unavoidable, our advocacy for them is generally inescapable, even in our most impersonal and professional engagements. In the social work arena, for instance, I think it fair to say that evaluators are generally sympathetic to the promotion of social welfare and the wellbeing of clients. But, of course, there are also 'welfare' areas where potential advocacies are not clear. Regarding the INTEGRA evaluation, for instance, programme clients, as benefit-claiming drug-using offenders, can provoke a range of reactions, often negative, from welfare professionals, as well as from the public and the media. 'Pathetic', 'drug-abusing', 'deviants' is a compilation of some singular descriptives from one researcher I know in the welfare arena. Advocating their interests within the enforcement and control imperative of a Probation Service initiative may seem controversial to some – although not to me.

Without any agreed guidelines as to what constitutes 'good' and 'bad' advocacy, the question is what should we do about it, if anything at all?

Jennifer Greene (1995) is unequivocal in her position that evaluators should actively advocate for the programme's target groups:

Evaluation inherently involves advocacy, so the important question becomes advocacy for whom. The most defensible answer to this question is that evaluation should advocate for the interests of program participants. (p.1)

Eleanor Chelmsky (1995), on the other hand, took the occasion of her 1995 Presidential address to the American Evaluation Association to warn against being perceived as taking sides:

What seems least well understood, in my judgement, is the dramatically negative and long-term impact on credibility of the appearance of advocacy in an evaluation. There is a vast cemetery out there of unused evaluation findings that have been loudly or quietly rejected because they did not 'seem' objective. In short, evaluators' survival in a political environment depends heavily on their credibility, as does the use of their findings in policy. (p.219)

Recognising and accepting such risks, but still preferring action to silence, Linda Mabry (1997a) sees this as an individual responsibility:

I am accepting advocacy as an individual responsibility, not a generally recognized obligation, and that my acceptance exposes me to potential censure from within and without my professional community. As a person, as an evaluator, I acknowledge that the advocacies I press can be wrong, but inaction can also be wrong. I cannot think it right merely to observe injustice or impoverishment or denial of opportunity...merely to report when there is reason to believe reports will be ignored or findings distorted...I choose not to omit, not to disengage, not silence. I expect to anguish over whether and how to commit myself, to be challenged to justify...to regret mistakes bitterly – and I do. (p.201)

For Michael Patton (1997), neither more or less advocacy is morally superior:

My own view, focused as always on utility, is that these different stances, indeed the whole continuum of evaluator activism, constitute options for discussion and negotiation with primary intended users. (p.125)

Herein lies one of my problems with Patton's *Utilization-Focused* perspective (mentioned in my Introduction to this Chapter). It is all a little too neat, a little too politically complacent in some respects, and it certainly lacks some of the uncertainty which is so clearly evident in Linda Mabry's account. The *primary intended users* to which Patton refers may well be the emergent source for an evaluator's advocacy (or ethical) dilemma and uncertainty, and may

well require evaluator action independent of them. More will be said of Utilisation-Focused evaluation a little later in this Chapter.

Ernest House – advocating for social justice and democracy

As for Ernest House, his option has been to develop extensive arguments for a reformist social justice purpose in evaluation. His notion of justice finds its most practical application in his arguments for a 'fair evaluation agreement' (House, 1980), and his broader arguments for evaluation ethics (House, 1993). Underpinning both of these is a move from the conventional ideal of objectivity to a commitment for fairness and impartiality. Responding to mainstream critics who argue that his fair evaluation agreement is biased towards the interests of the disadvantaged, he writes:

It seems to me that making certain the interests of the disadvantaged are represented and seriously considered is not being biased, though it is certainly more egalitarian than much current practice. (House, 1991, pp.241-242)

More recently, in collaboration with philosopher Kenneth Howe (House and Howe, 1999), and continuing his thesis that evaluation is value laden and that evaluators hold value commitments which shape their designs, House redirects his advocacy toward the pursuit of democracy. In this, he argues that evaluations should meet three explicit requirements:

- They should include all major stakeholder interests and views in some form. Inclusion is not tokenistic;
- They should allow for extensive dialogue so that stakeholder views and interests are authentic, as represented in the evaluation. Dialogue is not tokenistic;
- They should provide for sufficient deliberation so that valid conclusions can be arrived at, deliberation that uses the expertise of evaluators. Deliberation is not tokenistic.

Evaluators themselves should not ignore imbalances of power or pretend that dialogue about evaluation is open when it is not. To do so is simply to endorse the existing social and power arrangements implicitly and to evade professional responsibility. Evaluators, then:

...have a duty not to abuse their authority...Like physicians who are constrained by the value of promoting health, evaluators are constrained by the value of promoting democracy...Nor should evaluators play the role of neutral facilitators without regard to democratic consequences...they must be savvy negotiators, willing to engage in compromise. But, on pain of being mere functionaries doing the bidding of the powers that be, they must set limits as to how far compromise can go and be *uncompromising* about unwarranted, self-serving, and morally objectionable claims that stakeholders might advance. Doing only what is practical is not sufficient...they must make stands on moral-political fundamentals. And they must do so independent of stakeholders if necessary. (p.136)

As a set of guiding principles for moral-political advocacy, I think the above provides a *useful* and *pragmatic* foundation, and is something to which I generally could subscribe. However, House and Howe's notion of *democracy* requires some clarification, not least as most evaluators should already have some broad conception of democracy and their role within it, even if they do not make this understanding explicit. According to House and Howe, the received evaluation view of democracy is consistent with what they call *emotive democracy*:

In emotive democracy value claims are put forth by stakeholders and accepted at face value, as are preferences. A value is a value is a value, and value claims compete against one another in the public arena. The strength of this view is that it explains how American democracy seems to work currently. The weakness is that it seems to embrace the status quo. (p.48)

In contrast, they propose the notion of *deliberative democracy*, and identify the three requirements mentioned on the previous page for deliberative democratic evaluation, namely: inclusion, dialogue, and deliberation. These are not so cleanly distinguished, but affect and reinforce one another. The quality of deliberation is not separable from the quality of dialogue, which, in turn, affects whether inclusion, as opposed to mere tokenism, is achieved. The deliberative democratic view, then, aspires to bring participants into dialogue in fundamental ways, so that they can authentically

represent their own interests – and in that process they may also determine what their 'real' interests are.

What House and Howe offer is not a model that prescribes how to do an evaluation. Rather, they propose a value framework to help evaluators aspire to a set of defensible principles for enhancing inclusion, dialogue, and deliberation. In this way, evaluation should be able to develop deeper understanding of programme limitations and possibilities, especially for disadvantaged groups. Compromise, and being compromised, are unavoidable hazards. But there is no escaping this predicament by invoking the idea that values are unimportant or beyond the scope of evaluation:

Evaluation is as good or bad as the value framework that constrains it, in the same way that it is as good or bad as the research methodology it employs. In fact the two cannot be disentangled. (p.137)

In a similar vein to Scriven's Key Evaluation Checklist, House and Howe have produced a Deliberative Democratic Evaluation Checklist to help evaluators keep on track with their value framework.

House and Howe acknowledge that other evaluators advocate practices that are consistent with some of the views they endorse, including, what they call, *hyperegalitarian* relativist constructivists such as Guba and Lincoln (1989), as well as critical evaluators with a commitment towards emancipatory and participatory ideals. Indeed, they even recognise points of agreement with, what they call, *hyperpluralist* and *postdemocratic*, postmodernists:

Viewed in terms of avoiding authoritarianism (as opposed to *authoritativeness*), our differences with radical constructivist and postmodern approaches may not be irreconcilable. We are advised by them to tread lightly, to be tentative, and to be highly suspicious of those who claim to know what is best. We are also advised to pay close attention to local social conditions and to individual 'subjectivities'. We do not deny the legitimacy of these concerns. (House and Howe, 1999, p.87 – my interpolation)

Nevertheless, they consistently denounce radical constructivists and postmodernists as moral-political inactivists, that is, if they adhere to a philosophical constancy...

...A FEW WORDS ABOUT POSTMODERNISM (*sic*)

Consistent with its nature, postmodernism does not lend itself to handy definition. It rejects representation, and champions the poverty of language in conveying complex understanding. It is:

...the moment at which definitions begin to crumble
and the sign floats independently of the referent.
(Wakefield, 1990, p.20)

However, if shoved towards, an albeit contested sense of, philosophical 'purity', postmodernism can be caricatured as demanding silence:

There is no Truth to be spoken, no explanation without distortion, no theory without unacceptably coercive effect; even local and personal meanings invite interrogation. There is no valid representation of reality, reality being a construction and valid representation being precluded by linguistic indeterminacy, the remoteness of symbols from their objects, and unstated or unconscious motive. All is text, and all texts must be deconstructed to reveal inevitable inconsistencies, omissions, and inadvertencies. There is no legitimate author, authority denied; no legitimation for publishing, publication implying approval either from institutions complicit in oppressive power structures or from misguided alternatives which, given opportunity, would yield new oppressions. (Mabry, 1997b, p.1)

In its distilled extreme, then, postmodernism is so bleak, so sceptical, that few could probably tolerate it as a consistent personal philosophy. And certainly, it would be difficult to envisage a direct design relevance for the field of programme evaluation as I understand it, for not only does it disdain the status quo, it also disdains any strategies for improvement. How can postmodernists defend the practice of evaluation when the major thrust of postmodernism is to deny that such activities could be defensible? At first glance, it is difficult to disagree with Adelman:

Holding to postmodernism is incommensurate with being an evaluator. (Adelman, 1996, p.291)

Programme evaluations are generally characterised by some immediacy and specific focus. They are expected to address issues identified by, or negotiated with, stakeholders. Their scope is defined by contract and limited by resources. In many instances, a single study may need to describe the programme and its context, identify shortfalls between goals and outcomes, detect process pressure points, discover determinants and constraints, consider cost factors, and possibly specify some sense of overall merit. These are modernist demands, which presume the reality of programmes, the communicability of language, and the capacity for making decisions which aim to advance some social good – or at least aim to advance some agenda. In the stakeholders' world:

...there is little tolerance for questioning the linearity of time or intersubjectivity or rationality. The evaluator agrees to work in this world. (Mabry, 1997b, p.9)

The solution for postmodern evaluators, then, is generally to spin modifications. So, Mabry (1997b), for instance, draws a distinction, albeit not unambiguous, between extreme (sceptical) and moderate (affirmative) postmodernism:

- Sceptical postmodernism is that which has already been caricatured. It has given up on the notion of grand plans, on any rationality for devising them, and on any science or politics of any persuasion for achieving them. Instead, the entire focus is on the individual, and entirely on diversity rather than consensus.
- Affirmative postmodernism, on the other hand, rejects the unqualified relativism of the sceptical. While it retains emphasis on the individual and the unique, and a preference for intuitive interpretation and situated personal meaning over theory and grand explanation, it is less cynical and less restrictive: some behaviour is more justifiable than others, some values more worthy, some truths more true, some criteria for deciding among them more appropriate. Although truth cannot be known, and representations of it are likely to distort and conceal, the attempt to achieve clarity is worthwhile, and is enhanced by communication, however imperfect that is. Affirmative postmodernism, then, is not silent, and methodologically it resonates with constructivism.

It is distilled sceptical postmodernism which attracts the accusation from House and others of moral-political inactivism. Alternatively,

affirmative postmodernism is criticised for its philosophical inconsistency:

...if they do commit to moral-political principles to guide and evaluate practice, they can be (and are) charged with being inconsistent, although they might well counter with characteristic irony that intellectual inconsistency does not impede action. (House and Howe, 1999, p.74)

Indeed, is it postmodernism? Continuing a paradigmatic stance, is affirmative postmodernism not simply constructivism renamed? Or is it simply some notion of anti-modernist modernism? After all, as Cahoone puts it, 'modernity has been criticising itself all its life':

...postmodernism is the latest wave in the critique of the Enlightenment, the critique of the cultural principles characteristic of modern society that trace their legacy to the eighteenth century, a critique that has been going on since that time. (Cahoone, 1996, p2)

For me though, as an evaluator, and beyond scholarly or chic debate, I do not think any of this really matters - with one caveat. Owing the badge of 'postmodern evaluator' (currently) invites criticism, invites baggage, you can probably do without. This criticism may well come from an absolutist perspective, which specifically focuses on the extremes of relativism, and its general irreconcilability with the evaluation act. Certainly, this characterises House' main critique of postmodernism, and indeed constructivism. It is also characteristic of other realist appraisals, not least the analysis offered by Pawson and Tilley (1997) – see later.

So, rather than entertain any notion of *being* a postmodern evaluator, even though I share the constructivist thrust of such a perspective, I prefer to think of myself as someone who, I have to say, has moments (moods) when my constructivist hesitation is pulled even further in a sceptical direction:

At many other times, we will be modern, or schematic, or absolutist – the very things we bemoan in our postmodern mood. (Stake, 1997, p.43)

Instead of lamenting such sceptical moods, though, and whilst recognising their general incompatibility with direct evaluation

practice, I believe it could be useful to embrace them in a much more considered way. Sceptical postmodernism is intellectually compelling and, despite its variety of referents and some disunity, it presents internal consistency. I believe it has much to offer as a stance for reflexivity and meta-evaluation.

A Good Way to Think if a Bad Way to Act?

One practical implication of accepting the inevitable role of the evaluator in the evaluation process, as generally characterised by constructivist approaches, is that this could be highlighted or revealed in the evaluation report. Reinharz (1983) for instance, addressing this issue with regard to research generally, contrasts the conventional approach to conducting and reporting inquiry where the inquirers' attitudes are not revealed or analysed, with the alternative view that inquirers' attitudes should be fully described and discussed, and their values 'acknowledged, revealed and labelled' (1983, p.172). In a similar vein, Lincoln and Guba (1985) advise the keeping of a 'reflexive journal'. They suggest that this should include a personal diary where reflections are noted on the role of one's own values and interests. They see this as forming an important part of the study's 'paper-trail', and which could be open to external audit.

Latour (1988), on the other hand, expresses reservations about what he terms 'meta-reflexivity', the major function of which is to promote scepticism in a study's report. However, promoting such scepticism is precisely what I am proposing here. To be more precise, though, I am proposing that evaluators actively embrace and promote scepticism particularly when considering their reports *postevaluation*, or when considering the work of others. Unless specifically agreed with stakeholders, I would not wish any thrust for enhanced scepticism to risk potential use of any evaluation by unnecessarily over-obfuscating whatever value it holds (see next section).

Evaluation reports, as with research reports generally, are constructed texts, and, as such, full of subjectivities. The tendency, naturally enough from a utilisation perspective, is to present stakeholders with a tidy and digestible picture of the evaluation process and the data it generates. Being less generous about evaluator reporting motives, I would also argue that their texts may also be written from a self-promotional agenda, using subtle stylistic devices to persuade readers' perception of their evaluation competence. As MacLure (1999) argues, however, doing any of this merely masks the complexities and untidiness in our work, and

invites us to collude in a 'critical illiteracy' through which we ask no further questions. These further questions, though, might well prove a handy source for evaluator reflexivity, as well as useful meta-evaluation fodder.

Postmodernism challenges us to strip away the veneer of tidiness and 'deconstruct' reports to acknowledge and consider the assumptions, preconceptions, and values which underpin them:

Deconstruction seeks out the singular instances of failure in method, methodology, and philosophy...Such deconstruction does not seek to destroy but to identify and appreciate necessary failure. (Stronach, 1997, p. 35)

And,

...deconstruction *is* education, a necessary unsettling of what it is that we can ever claim to understand about the nature of the individual and the social. (Derrida, cited in Sohm, 1994, pp.28-41)

But, deconstruction is not:

...irrationalist or nihilistic: reasons must be given. (Bennington, 1996, p. 140).

For me, then, and within this context here, deconstruction offers a stance for evaluators to open for critical reflection and learning those things that evaluation reports tend to lock inside tidy and enduring pragmatic boxes.

With the INTEGRA report in mind, for instance, interesting deconstruction 'sites' might include: identity, empowerment, exclusion, change, contact, and validity.⁶ Fuller consideration, though, is now beyond the scope of this thesis. However, whilst not a method recipe, 'avoiding predictable reduction', **Stronach and MacLure's** (1997) *Educational Research Undone: The Postmodern Embrace*, offers a window, if albeit 'intentionally eccentric' (p.3), onto the inventive dance of deconstruction.

⁶ My original intention with this thesis had been to undertake a deconstruction of the INTEGRA evaluation report, concentrating on suitable sites related to the, generally, positivistic original core evaluation design of my manager, and the, generally, constructivist perspective of my contributions – the pseudopodia of my amoeba metaphor (see Chapter 4). I was advised against this path by my PhD supervisors.

ALTERNATIVE UTILISATION SHOOTS

Generally, paradigmatic alternatives to the conventional view of evaluation (as epitomised by Campbell and Scriven) can be seen as driven by a desire to produce deeper and *better* knowledge. However, the 1970s also witnessed the dawn of focused attempts to produce more *useful* knowledge. Great diversity was generated, with different theorists suggesting different ways of obtaining useful knowledge, and looking to different stakeholders to whom it might be useful. The issue of use, though, was not new to the field. In the earliest programme evaluation text, Suchman (1967) wrote:

The 'success' of an evaluation project will be largely dependent upon its usefulness to the administrator in improving services...Unlike the basic researcher, the applied researcher must be constantly aware of the potential utility of his findings. (p.21)

But the early theorists did not discuss use in detail. In their comprehensive assessment of evaluation theories and theorists, Shadish, Cook, and Leviton (1991) describe those theorists who sought to move from knowledge-driven to use-led emphases as 'Stage Two' theorists (p.171). They outline the work of three US theorists as epitomising this 'stage': Carol Weiss, Joseph Wholey, and Robert Stake (see earlier). The three authors suggest different ways of obtaining 'good use' and identify different stakeholders:

- Carol Weiss' key contribution to evaluation theory lies in her emphasis of the political context of policy research. She prioritises programme decision-makers and policy-makers at the federal and state level as the main users of evaluations. As I will expand a little later, she holds a longer-term view of evaluation use, and, generally, advocates 'knowledge creep and decision accretion' (Weiss, 1980), which lead to 'enlightenment', rather than 'instrumental' use of evaluation.
- Joseph Wholey (e.g. 1985), on the other hand, generally promotes 'instrumental' use among middle-level bureaucrats who are willing to manage for results and improve existing programmes in the short-term.
- Robert Stake, as previously discussed, believes that creating local solutions to local problems is best left to local stakeholders with local understanding of their programme, rather than evaluators or policy-makers in a distant bureaucracy. Like Weiss, he favours an 'enlightenment' approach to evaluation use.

Regardless of their differing foci, all three agree that evaluators should explicitly identify the users in each evaluation, and work closely with them in order to create useful information. In this, they supplant the 'Stage One' (Shadish, Cook, and Leviton, 1991, p.69) Campbell-Scriven image of evaluators as 'objective outsiders'.

In tackling the remainder of this section, I will reflect a little more on Weiss, but substitute Wholey and Stake with a more detailed consideration of Patton's instrumental *Utilization-Focused* perspective (1978, 1986, and 1997), a perspective developed, generally, from his work with community initiatives at the local level, and a perspective which informed much of my stance in relation to the INTEGRA evaluation:

'This focus derives from my work with small, community-based programs where the idea of conducting 'research' may be intimidating...' (Patton, 1997, p25)

Carol Weiss – linking policy research to evaluation

Thanks in large part to sociologist Carol Weiss, it is now widely recognised that politics and evaluation are joined at the hip. She set about rethinking the place of evaluation in policy-making by both explicating its political context and undertaking empirical work on ways in which evaluation is actually used.

The early Weiss first favoured instrumental use, where she expected that responsible policy-makers would implement answers provided by evaluation into policy or practice:

...evaluation assists decision-makers to make wise choices among future courses of action. Careful and unbiased data on the consequences of programs should improve decision-making. (Weiss, 1973, p.37)

However, the perception that 'evaluative data seemed to have little effect on either budgetary allocations or the selection of programs for expansion or reduction' (1987b, p.42) led her to rethink the place of evaluation in policy:

Evaluation is a rational enterprise that takes place in a political context. Political considerations intrude in three major ways, and the evaluator who fails to recognize their presence is in for a series of shocks and frustrations. (1973, p.37)

She identified these three intrusions:

- One stems from the fact that programmes 'are the creatures of political decisions. They...remain subject to pressures – both supportive and hostile – that arise out of the play of politics' (p.37). Programmes develop loyal constituencies who fight for programme survival, concerned more with their personal interests, such as career concerns and jobs, than whether the programme achieves its goals. Likewise, opponents have similar personal interests. 'Devastating evidence of program failure has left some policies or programs unscathed, and positive evidence has not shielded others from dissolution. Clearly, other factors weigh heavily in the politics of the decision process' (p.40).
- Second, evaluation entails political considerations because its reports enter the political arena. 'There, evaluative evidence of program outcomes has to compete for attention with other factors that carry weight in the political process' (p.37), and particularly ideology and interests (Weiss, 1983). Her subsequent empirical work focused on this area – see below after bullets.
- Third, 'evaluation itself has a political stance. By its very nature, it makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeability of others, the legitimacy of program goals and program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation' (Weiss, 1973, p.37). By accepting a programme emphasis on service delivery, most evaluations tend to ignore the social and institutional structures within which the problems of target groups are generated and sustained. Her general conclusion was to lament that 'it does appear that evaluation research is most likely to affect decisions when the researcher accepts the values, assumptions, and objectives of the decision-maker' (p.41).

Along with her colleagues, Weiss interviewed 155 senior officials in federal, state, and local mental health agencies. She found that officials and staff used research to provide information about service needs, evidence about what works, and to keep up with the field. However, it was also used as overlay, to legitimise positions, and to provide assurance that a position already held was the correct one. She concludes:

It was one source among many, and not usually powerful enough to drive the decision process. (Weiss, 1980, p.390)

And as for direct utilisation:

Instrumental use seems in fact to be rare, particularly when the issues are complex, the consequences are uncertain, and a multitude of actors are engaged in the decision-making process, i.e., in the making of policy. (1980, p.397)

Research use was also reflected in officials' views of the decision-making process. Decisions were seen as fragmented both vertically and horizontally within organisations, and were seen to be the result of gradual and amorphous steps:

A salient reason why they do not report the use of research for specific decisions is that many of them *do not believe that they make decisions*. (1980, p.398)

Policy is not the result of stop-go decisions about alternatives. Rather, it emerges from custom and implicit rules about what can be done, improvisation in new situations, mutual adjustment to the actions of other policy-makers, negotiation of conflicts, countermoves when bargaining breaks down, opportunity, and as a by-product of other decisions (Weiss, 1982). This process is better termed 'decision accretion' rather than decision-making.

Even if policy-makers had more leverage, they are:

...very busy people. More issues come at them each day than they have time to consider very carefully. (Weiss, 1987, p.275)

According to Weiss, policy-makers most value data that comes naturally to them, not data they have to work to obtain, and this includes media reports. They also value data provided by lobbies and interest groups, along with information they get from 'issue networks', such as those comprising academics and think tanks. Weiss urges evaluators to learn how to use these windows of opportunity for their work. Policymakers, she contends,:

...screen the research that they read or hear about through a series of implicit filters. If it fails to pass the filters, they tend to discard it. If it passes, it deposits a residue that becomes incorporated into their stock of knowledge, which they draw upon when action is called for. (Weiss and Bucuvalas, 1980, p.249)

Weiss categorises these implicit filters as 'Truth Tests', which include the perception made by policy-makers of the research quality, and 'Utility Tests', whether the research gives explicit guidance for feasible reform, or whether more fundamental change, with more consequent uncertainty and risk, is suggested (Weiss and Bucuvalas, 1980, pp.253-256).

Given such a diffuse process of policy-making, then, Weiss also proposes a diffuse approach to understanding evaluation use at the policy level. She suggests 'enlightenment' or 'knowledge creep' as being more valid than any rational or instrumental theory of use:

The enlightenment model of research...implies that research need not necessarily be geared to the operating feasibilities of today, but that research provides the intellectual background of concepts, orientations, and empirical generalizations that inform policy. As new concepts and data emerge, their gradual cumulative effect can be to change the conventions policymakers abide by and to reorder the goals and priorities of the practical policy world. (Weiss, 1977, p.544)

After years of further research, Weiss concludes that evaluations seldom determine the outcome of policy-making. However, she still believes that policy will be better served if officials pay more attention to evaluation results, and she urges evaluators to double their efforts to get their message across:

Even if we realise that evaluation is not the star in the policy drama, we have a responsibility to communicate the best information and analysis available to the principal players. (Weiss, 1999, p. 483).

Weiss' position has been challenged, especially by Michael Patton who I will consider in the next section. His argument is twofold. First,

he argues that Weiss wrongly generalises from policy research to programme evaluation:

It makes sense that policy research would be used in more diffuse and less direct ways than program evaluation...(because they are)...different kinds and levels of practice. (Patton, 1988, p.12 – my interpolation)

I agree with Patton on this point. Weiss' work has been undertaken with policy-makers at the US Federal and State level who are, generally, responsible for broad-based social programme planning. Patton's work on the other hand, and which takes an instrumental stance in relation to use, occurs at the, generally, local level. Patton tends to work with stakeholders who are not so enmeshed in a broad public policy framework, and are therefore more potentially able to use evaluation results to make local policy and programme decisions. It should also be noted that Weiss, according to Shadish, Cook, and Leviton (1991 p.190):

after the mid 1970s...rarely discusses 'evaluation', mostly discussing policy research or social science research. This...might lead readers to think she abandoned evaluation. This is not so. Weiss simply stopped making distinctions among program evaluation, policy research, and applied social science.

Patton's second concern with Weiss is that her position seems 'quite dismal': 'The Weiss vision, in my judgement, is not marketable' (Patton, 1988, p.11).

On this point I have to disagree with Patton. Although theories of enlightenment defy comfortable demonstration, they do seem plausible. They also find congruence in important strands of British policy research, and I will return to aspects of this in Chapter 3.

As for Weiss' response to Patton, she complains that in his world 'everybody behaves rationally' (Weiss, 1988, p.18), and that he underestimates the political factors which make instrumental use more problematic:

I doubt that we can ever persuade stakeholders to make evaluation results the overriding consideration in program decisions. For one thing, program people know a lot more about their programs than simply the things the evaluator tells them. They have firsthand experience in the operating organization; they know the site, the clients, the staff, the problems, the budgets, the conflicting directives from sponsors and funders, the state of relationships with other organisations that refer or receive clients, the history, the complaints and kudos, and the prospects for the future. (p.17)

I have some sympathy with Weiss on this point. There are some aspects of Patton which, for me too, do feel politically complacent. However, the point is surely not to expect guarantees of instrumental use, but to strive for a practice framework that makes it more possible. And in this, I find Patton's work highly instructive. It should also be noted that Weiss' 1988 response to Patton is based (I assume) on his 1986 2nd edition of *Utilization-Focused Evaluation*. As Patton acknowledges in the Preface to his 1997 3rd edition:

The second edition...was alternatively brash and shy, assertive and uncertain, like an adolescent coming of age...I wanted to set the record straight and clarify points of confusion. By my own criteria, I only partially succeeded... (Patton, 1997, p.xiv)

Michael Patton – more than the Lewis Carroll of evaluators?

In their evaluation 'think piece', *Realistic Evaluation*, Pawson and Tilley (1997) dismiss Patton (with characteristic disdain I feel) as the 'Lewis Carroll of evaluators', who...:

...uses every analogy, tale and metaphor in the book...to promote a more skilful approach to evaluation...we fear there is one old adage he has not come upon...'metaphors are no arguments, my pretty maiden'...whilst it would be nice if we could get away with some fundamentally English metaphors in evaluation training (bowl a good line and length, play with a straight bat etc) we suppose that matters of research design and data construction need to be rooted in a very clear-headed understanding of social change and social explanation. (p.16)

Not only do they miss the point, but they misrepresent it too. However, leaving more detailed consideration of Pawson and Tilley for a little later, I have to say I personally find Michael Quinn Patton a cracking communicator, something which I feel may well stem from his long, extensive, and varied evaluation practice. I also feel that Patton's framework for promoting evaluation use has probably become the most influential evaluation text in the past ten years. I have no evidence for claiming this, save my interpretation of evaluation literature and online searching – but it would be interesting to internationalise and update Shadish and Epstein's (1987) US survey of evaluators' values, practices, and influences (see footnote 4) to ponder how defensible this claim may be.

Essentially, *Utilization-Focused Evaluation* (1997) seeks to ensure that the conduct of an evaluation, and its results, are relevant to the questions asked by those who have a stake in the programme, and is useful in so far as it can effect positive change. It is underpinned by the premise that:

...evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, *from beginning to end*, will affect use' (Patton, 1997, p. 20).

Nor is use an abstraction:

Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. (p.20)

In any evaluation, there are many potential stakeholders with varying degrees of interest in the programme. A key initial task for the evaluator, then, lies in undertaking a stakeholder analysis to identify the *primary intended users*, and then working with them to further identify the *primary intended uses*:

Since no evaluation can be value free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to apply evaluation findings and implement recommendations. (p.21)

In this way, utilisation-focused evaluation is intended to be highly personal and situational. The evaluator's job in the early stages is to facilitate a working relationship with intended users to help them consider what kind of evaluation questions they want to ask, or reconsider the evaluation questions they already have in mind, and then determine what evaluation design would be appropriate. Where possible and practical:

An evaluation task force can be organised to make major decisions about the focus, methods, and purpose of the evaluation. The task force is a vehicle for actively involving key stakeholders in the evaluation. (p.353)

Underpinning the task force is the idea that intended users are more likely to use evaluations if they understand, and feel ownership of, the evaluation process and results:

...by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way. (p.22)

Characterising all phases of evaluator-user interactions is the prescriptive and descriptive notion of 'active-reactive-adaptive'. This describes how decision-making unfolds, but is also prescriptive in alerting evaluators to consciously act, react, and adapt in order to increase the effectiveness of their work with stakeholders:

Utilization-focused evaluators are, first of all, active in deliberately and calculatedly identifying intended users and focusing useful questions. They are reactive in listening to intended users and responding to what they learn about the particular situation in which the evaluation unfolds. They are adaptive in altering evaluation questions and designs in light of their increased understanding of the situation and changing conditions. (p.135)

In this way, the evaluation, and the evaluator's role, can be seen to be organic. Hence my use of amoeba as a metaphor for the INTEGRA evaluation, a metaphor I used to positive effect with the INTEGRA Project Team to explore with them their sense of the evaluation as it unfolded (see Chapter 4). Being active-reactive-

adaptive, being situationally-responsive, then, means that evaluators do not impose 'cook-book' designs, evaluators:

...don't do the same thing time after time. They are genuinely immersed in the challenges of each new setting and authentically responsive to the intended users of each new evaluation. (p.135)

Their menu of choices embraces the full range of evaluation methods and evaluator roles, depending on the negotiated questions to be addressed, the stage of the programme life cycle, the evaluator's location (internal employee or external consultant), and:

the evaluator's own personal knowledge, skills, style, values and ethics. (p.136)

This utilisation-focused commitment to situational responsiveness is controversial and attracts some criticism within the field. Scriven, for instance, as we have seen earlier, advocates only one role for the evaluator – namely that the evaluator renders 'distanced' judgement about merit or worth. Pawson and Tilley (1997), in dismissing Patton as 'the Lewis Carroll of evaluators' argue:

Evoking 'skill' in method is never enough; it can never tell us when and why to utilize a particular approach. (p.16)

I believe their criticism is unfair. Patton does not aim to provide a methods textbook. Rather he outlines a philosophy, with practice principles, to guide action within a utilisation-focused framework. That said, Pawson and Tilley do have a defensible point to make:

If the contract is, say, for 'an evaluation' of the efficacy of a jelly diet on prisoner rehabilitation, the researcher can indeed use a full repertoire of skills...suppose, however, the decision maker's wants are expressed at the level of a preference for a particular methodological strategy, for example 'a quasi-experimental investigation' of the outcomes of a jelly diet...research skills would still...be applied in doing the deed but other strategic considerations would

have been foreclosed, giving us a general rule of pragmatic evaluation: *the more explicit the policy mandate, the more compressed and purely technical the researcher's role.* (p.17)

However, as Patton makes clear: 'a utilization-focused evaluator is not passive in simply accepting and buying into whatever an intended user initially desires' (Patton, 1997, p.364). Being active connotes an obligation on the part of the evaluator to represent the standards and principles of professional competence and ethical behaviour, as well as their own sense of morality and integrity. Pawson and Tilley's concern that 'he who pays the piper *might* call the tune' is valid, but this does not mean that 'he who pays the piper *does* call the tune'. I believe Patton in the 90s is more explicit in outlining such political and ethical concerns. Perhaps that is why Pawson and Tilley select a 1982 quote to denote his stance regarding this:

If a funding mandate calls for a summative outcomes evaluation, then the evaluator had better be prepared to produce such an animal, complete with a final report that includes that terminology right there on the front page, in big letters, in the title. (Patton, 1982, p.49)

And I certainly believe they are being somewhat disingenuous with their critique of Patton when they continue with their 'jelly diet' scenario:

The ultimate squeeze, of course, is when the policy maker demands 'evaluation results': show us that the jelly diet works! Here we have the Rothschildian vision with research skills for hire – skills, moreover, which are closer in esteem to those of the caterer or plumber than those of the scientist. (p.17)

However, notwithstanding Pawson and Tilley's (apparent) views concerning the skill value of caterers and plumbers, and (presumably) their willingness to supply substandard sandwiches or stopcocks, the issue of ethics in utilisation-focused evaluation is highly pertinent and will be covered further in Chapter 5.

My personal reservations with Patton generally concern the seeming political tidiness and complacency of his framework, and aspects of the pluralist pragmatism it exudes. The first point has

been touched on previously, particularly with regard to advocacy dilemmas and instrumental evaluation use at the broad policy level. The second point though, whilst linked to the first, requires some clarification.

A Few Words About Pragmatism

Philosophical pragmatism draws particularly on the work of late 19th century and early 20th century American philosophers such as Pierce, James, and Dewey, but has more recently been reinvigorated by, amongst others, Richard Rorty. In essence, he proposes that we should abandon epistemology. In Rorty's 'postepistemological' thinking, the history of philosophy shows that there are no final answers to the traditional questions about 'knowledge', 'truth', and 'representation' (Rorty, 1991), and so, consequently, they should be rejected. His arguments have exercised a growing influence in evaluation.

At this point, though, it is worth distinguishing between the philosophical pragmatism of Rorty, and the *methodological* pragmatism such thinking, in part, has brought to programme evaluation:

Methodological pragmatism rests on an impatience with philosophy and an emphasis on real world evaluation and practice, and it is claimed that methods can be separated from the epistemology out of which they emerged. The emphasis is thus on practical utility and the credibility of the methods used. (Shaw, 1999, p.51)

Such impatience with philosophy in evaluation owes much to a corresponding impatience with the 'paradigms debate' which has paralleled the development of evaluation as a discipline. Throughout the 1970s and 80s, voices from the poles of the two-paradigm positivist/constructivist evaluation world outlined earlier noisily defended their respective positions as the right and righteous path to 'truth'. Thomas Cook (1997), one of the principal US actors in the reform of experimental approaches to programme evaluation, captures a common understanding when he argues that the debate was principally about the legitimacy of qualitative approaches. As this has now been achieved, it is time to move on towards better evaluation theory:

The debate has been very helpful, and the qualitative advocates have deservedly won full equality in today's evaluation enterprise. However, the debate is now diverting intellectual energy from more urgent tasks, and we need to move on. (Cook, 1997, p.31)

According to the current mainstream position, then, the:

...debate has run out of intellectual steam and is now relegated to comedy on the internet. (Patton, 1997, p.266)

And mainstream thinking is neatly caricatured by Patton's notion of a 'Paradigm of Choices':

The focus has shifted to methodological appropriateness rather than orthodoxy, methodological creativity rather than rigid adherence to a paradigm, and methodological flexibility rather than conformity to a narrow set of rules...The debate over which paradigm was the right path to *truth* has been replaced, at the level of methods, by a paradigm of choices. (p.295, p.298)

The connection between philosophical paradigms and methods has, apparently, been broken. Whilst I personally understand the desire to move on from the noise of the old paradigms debate, I believe it is naïve to believe that evaluation design and methodology can be separated from philosophy in such a way. As outlined in Chapter 1, the question today is not so much about which paradigm offers the best way to best truth, but whether, and to what extent, philosophy matters. This is a question not only for evaluators in considering what they bring to any evaluation and its reporting, but also for stakeholders in their understanding of, in what meaning(s) they take and use from, the evaluation findings. In this, I believe philosophical challenges to the current pragmatic paradigm of choices will bear fruit in reinvigorating aspects of the debate, and help in the further development of programme evaluation theory. As a way of concluding this speed-mapping of the evaluation landscape, I would like to outline one recent UK development, which, whilst sharing a commitment to move from the noise of the old paradigms debate, has its philosophical roots firmly embedded in postpositivism.

BEYOND THE PARADIGM OF CHOICES?

This development is particularly relevant for this thesis as it is beginning to have a significant impact on evaluation thinking within the UK Criminal Justice System:

Ray Pawson and Nick Tilley – Realistic Evaluation

At the outset, I ought to say, and as may be evident from comments I have made about Pawson and Tilley's (1997) *Realistic Evaluation* in preceding pages, there are aspects of this book that I find personally distasteful. Not least is, what I read as, a lack of an even hand in reviewing the work of other evaluators, from whichever paradigm perspective (other than their own scientific realism) they may generally be writing. For academics who proclaim a commitment to 'scientific endeavour', I find this disconcerting. However, and from my own media experience, I also recognise the publishing pressures which can be exerted on authors to make their work more commercially attractive and readable. Hullabaloo and humour can help in this, as can stylistic use of 'superb' superlatives in promoting one's own stance.

That said, I believe the work of Pawson and Tilley represents a significant contribution to the development of evaluation theory, and is already gaining particular credence within the UK Criminal Justice System, as evidenced by Merrington and Hine's (2001) adoption of it as their model of choice, albeit within Patton's utilisation-focused commitment, in their Probation Service evaluation handbook:

...(Pawson and Tilley) argue that it is not enough to know 'What Works', and that the question that policy makers and service delivery staff need to address is 'What works, for whom, in what circumstances?' This handbook supports this model, which uniquely emphasises the role of theory and the development of theory as a means of improving understanding. Pawson and Tilley argue that evaluators need to think through the whole process of why a programme might work, including the inputs, the outputs and the context in which the work takes place. They argue that it is the interrelationship between the factors that creates mechanisms for change and that, without understanding the nature of the mechanisms, it is very difficult to understand why a programme is successful or not. (Merrington and Hine, 2001, p.2.9)

Essentially, Pawson and Tilley seek to incorporate the critiques of positivism without abandoning the concept of a knowable reality. The roots of their approach lie in the realist traditions of the European philosophy of science, as echoed in the writings of, amongst others, Harré (1972), and Bhaskar (1997). Within the evaluation field, their work can be seen as a further refinement of the postpositivist critical realism generally epitomised by Campbell and Scriven. In this sense, reality exists external to the knower, and although it cannot be fully apprehended, we can strive for a continually-improving approximation of it. However, scientific realism goes further in understanding the world as a fluid open system, as a constellation of structures, mechanisms and contexts. In attempting to draw parallels here with social programmes, Pawson and Tilley invoke four realist concepts:

- **Embeddedness** - of all human action within a wider range of social processes. This is the 'stratified nature of social reality'. Actions make sense only because they contain in-built assumptions about a wider set of social rules and institutions. So, for example, signing a check is routinely accepted as payment, but only because we accept its place within the social system known as banking. Explaining actions in terms of their location within different layers of social reality results in a rejection of the 'successionist' view of causation i.e. as a relationship between discrete events (cause and effect). Rather, realists embrace a 'generative' sense of causation which relies on the concept of...
- **Explanatory Mechanism** – this is not a variable, but an account of the nature, behaviour, and interrelationships of those processes that are responsible for an outcome. This includes stakeholder reasoning and choices, and their capacity to put these into practice. A mechanism, then, is a theory. Through the notion of programme mechanisms, we can move from asking whether a programme works, to considering what it is about a programme that makes it work.
- **Context** – is the partner concept of mechanism. The relationship between causal mechanisms and their effects is not fixed, but contingent. All social programmes engage with prevailing contextual conditions. These include the social rules, relations, and cultural/political systems that frame the programme mechanisms. A central task of realistic evaluation is to investigate the extent to which such context(s) enable or disable the intended mechanism of change.

- **Outcomes** – the objective of realist inquiry is to explain outcomes, or outcome patterns.

Evaluations produce context-mechanism-outcome configurations (CMOCs). The initial stage of realistic evaluation, then, involves developing some startpoint CMOC theories or models. These can come from various sources, such as: social science theory, results of previous evaluations, discussions with stakeholders, and tacit knowledge. The empirical part of the evaluation next comprises a multi-method exploration of these CMOCs, addressing the question of what actually works, for whom, and in what contexts. This then feeds back into the starting theory in a continuous cycle:

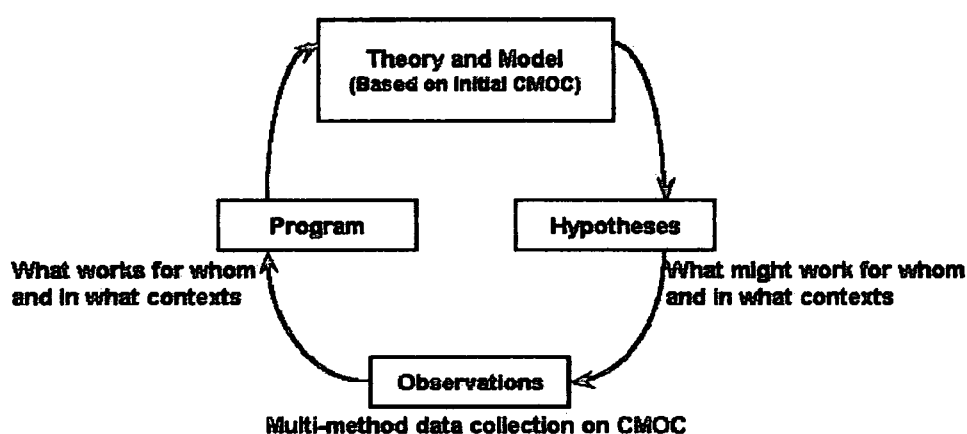


Fig. 2.1 The Realist Evaluation Cycle
(Adapted from Pawson and Tilley, 1997, p.85)

In this way, realistic evaluation aims to offer a holistic perspective, and sets out to address the complexity of programmes, including their process and the context(s) in which they operate. As such, it does not dismiss the potential contributions of any methodology in data gathering. Does this mean that Pawson and Tilley see themselves as signed-up members of the pragmatic pluralist party?:

Our answer is 'no', and indeed on several counts 'no!'. Our initial hesitation stems from the fact that pluralism, with its 'a bit of this and a bit of that' approach, actually leads to no new thinking beyond the ill-defined compromise... Secondly, it is woefully weak in knowing where to stop. That is to say, as information and potential stakeholders begin to multiply under

pluralism, we have no clear guidelines on whose views to prioritize, on what to do if viewpoints differ, on whether we want attitudinal or behavioural information from each group, and so on. (Pawson and Tilley, 1997, pp.154-155)

Rather, what is important is that any selection of methods and any choice of respondents, along with any selection of instruments or interview approaches and questions to be addressed, is driven by the developing theory. The realist approach can thus be seen as a family member of the broader 'theory-driven' perspectives of evaluation. All theory-driven approaches begin with the premise that every social programme is based on theory, some idea about how and why it will work. This may be explicit or implicit. The key to understanding what really matters about the programme is through identifying this theory (Chen, 1990; Weiss, 1995). This process is also known as developing a programme logic model, describing how the program works (In Chapter 3, I construct just such a programme logic model for the INTEGRA Project). The family of theory-driven approaches, though, is generally characterised by disagreements about the exact scope and content of the theory required. What distinguishes realistic evaluation is that it delivers a prescription for theory development based on the explanatory ingredients of mechanisms, contexts, and outcomes (In Chapter 6 I construct a refined model for the INTEGRA Project based on the findings from the evaluation, including process and context data which emerged during the evaluation process).

To summarise, then, the central purpose of realistic evaluation is to improve the programme, both in terms of its content and targeting, to improve the theory, and to improve the multi-method mix of data gathering techniques, in a continuous cycle of improvement. As programme effectiveness can be subverted or enhanced through the unanticipated intrusion of new contexts and new causal powers, the model cannot be seen as a static entity which is tested in a period of time and subsequently judged on its success. If it is going to be effective in a generative sense, it has to change in line with the changing and permeable social world. In all of this, the ultimate aim is to gradually lay down 'outcome typologies in which the successes of whole program families are explained' (Pawson and Tilley, 1997, p.86).

There is much in the work of Pawson and Tilley that I find personally appealing – even if I dislike aspects of their writing attitude. Their emphasis on process, context and human agency, their capacity for

an inclusive participatory commitment to theory-building and testing, the space they allow for emergence, and even their realist notion of a stratified social reality, do not offend my constructivist or social justice sensitivities. There is much in their perspective, I am sure, that has underpinned, albeit in an intuitive and utopian sense, some of my past specialist roles in developing practice for work with drug users. Indeed, Mansoor Kazi (2000), Director of the Centre for Evaluation Studies at Huddersfield University, applies a scientific realist perspective to his work in undertaking evidence-based approaches to social work practice development.

The problem for me in the past, though, and as it is now, is grasping how to devote sufficient resources to make it happen with integrity, and how to do it in such a data-gathering way that the inquiry itself hand-in-gloves the practice, the programme. I can appreciate, for instance, how realistic evaluation could be applied to well-resourced Probation Service Pathfinder⁷ pilot projects, where evaluation is central to the programme remit. However, I am not convinced that this is an approach which would be suitable, at least if undertaken in a holistically-integrative way, in many other instances. As such, the extent to which programme operation is actually approximated will depend on the pragmatic decisions that are made at the outset regarding the feasibility of the proposed design. Just how inclusive of less powerful voices would the theory-building and testing be? How contextually-inclusive of more powerful sharp-edged structures would the theory be? What compromises might be made in relation to multi-method data-gathering and analysis? How many laps of the realist evaluation cycle would be undertaken? With all of these questions, I would worry that many fine claims in the name of realistic evaluation could well be made from less than fine applications of the approach.

WHAT IS EVALUATION? PART 2

In the introduction to this Chapter, I undertook to consider the question of *What is evaluation?* in two parts. My initial exploration earlier, in *What Is Evaluation? Part 1*, encapsulates the partial consideration that carried me through my participation in the INTEGRA evaluation. Much of this, as should be evident from Chapter 5, was driven by Patton's utilisation-focused perspective,

⁷ Pathfinders are an integral component of the 'What Works' initiative. Essentially they are 'pilot schemes' funded at the local Probation service level. Those that are shown, after competent evaluation, to deliver positive results with offenders receive formal accreditation before being made available nationally.

along with practice-derived intuition, a very basic sense of social research methods, and unarticulated constructivist hesitation. My answer here, though, is somewhat different. It represents, what I feel to be, an enriched understanding informed by the additional thinking I needed to tackle this Chapter. Or, perhaps I should say it represents an enriched confusion.

The field of evaluation is immense and complex. In trying to map some of its scale, both philosophical and methodological, and in trying to situate my personal stance in relation to it all, I have chosen to focus on selected key evaluation figures. This might seem a risky business. My linchpin could well be someone else's spare part. However, my choices have been driven by a purpose. And that was to lead the way to this section, and specifically to this question: *what is evaluation?* And to answer that question now, without further prevarication: evaluation, for me, put simply, is a personal construct, and, as such, and unlike research, it resists generalisation. Each of the key evaluation theorists outlined carries some ingredient which constitutes a part of my (current) construct. Before I proceed to more detail, I need to identify one evaluation theorist, listed only in passing so far, whose ideas have had the biggest impact on my thinking more than any other, and whose work undergirds my approach to the writing of this Chapter and this thesis: **Saville Kushner**.

In *Personalizing Evaluation* (Kushner, 2000), as well as explicating his thought-provoking blend of democratic and participatory evaluation approaches, Kushner also seeks to personalise the role of evaluator, and construct this as a form of personal and political expression:

We need to be able to think independently of methodology so as to use it as a site of reconciliation between our personal values as evaluators and the volatile characteristics of the field we are observing – and that ought to be part of the professional preparation of young researchers and doctoral students. We need to make methodology a striving to combine justice with intellectual autonomy – and to some extent this requires us to find our own 'voices' in enquiry, as well as continuing the search for better ways of giving voice to those we evaluate and acknowledging the changing complexities of their lives. (Kushner, 2000, pp.14-15)

So, what do each of this Chapter's selected theorists bring to my personal construct, and the overarching purpose, of programme evaluation? In the order in which they appear:

- Campbell – quantitative deduction and fallible realism;
- Scriven – summative vs formative, and evaluating evaluation;
- Stake – qualitative discovery and constructivist hesitation;
- Guba and Lincoln – context and stakeholder constructs;
- House – advocacy, social justice, ethics, and democracy;
- Stronach and MacLure – deconstruction for reflexivity;
- Weiss – policy 'knowledge creep' and enlightenment;
- Patton – local instrumental use and situational responsiveness;
- Pawson and Tilley – multi-method, holistic, and theory;
- Kushner – this is personal.

Even if I do not necessarily agree with their theoretical justifications, there are design and methodological practices in the work of each of these theorists that I would endorse. Good practice, I believe, is eclectic and informed by theory, not puristically derived from it. Just as a narrow focus on experimental and quasi-experimental approaches proved insufficient to the task in evaluation's early years, so today a narrow focus on any one other approach is equally insufficient to the complexities of the field. A credible evaluation construct needs a place for multiple approaches, and must be clear about the contingencies which guide the selection of one over another when programme circumstance, stakeholder requirements, and resources force that choice.

One such contingency is philosophical preference or prejudice. I believe methodological pragmatists are naïve in arguing that the bonds between philosophy and method have been broken. But I am not pushing for some notion of paradigm *purity*. Rather, I would support the view of Greene when she argues of epistemological *integrity* in research:

...epistemological integrity does get meaningful research done right. (Greene, 1990, p.229)

People's actions and decisions in evaluation, as in any other activity, are shaped by values and worldviews. I believe it behoves each of us, if we strive for any sense of evaluation professionalism, to explore and articulate our philosophical perspectives and be alert to how these impact on our work. As outlined in Chapter 1, one of the aims of this thesis is to undertake just such an exploration with

regard to my role in, and contributions made to, the INTEGRA evaluation.

As should be clear from this Chapter, I understand myself to have a philosophical centre of gravity generally depicted by the constructivist paradigm. Hence my desire throughout the INTEGRA evaluation to inject qualitative and formative components (see Chapter 4). However, I do not operate at the caricatured pole, and do not hold to relativism so tightly. Indeed, there are aspects of realism that I find appealing, particularly the scientific realism underpinning the work of Pawson and Tilley (1997). Perhaps this betrays Popperian remnants of my biological science roots? I find it interesting, for instance, that I choose a biology metaphor, the amoeba, to 'model' my sense of the INTEGRA evaluation (Chapter 4). Into this mix I would also add aspects of critical evaluation as denoted by House and Howe (1999). And underpinning all of this, a teleological commitment to stakeholder-focused instrumental use as the local purpose of evaluation, and enlightenment at the macro policy level.

This last point can serve to illustrate a site of further complexity. Constructs are not static. Writing this Chapter, for instance, has prompted particular thinking around the purpose of evaluation. My strong sense of 'use' as the Holy Grail of evaluative purpose begs interrogation. Indeed, is 'use' a sufficiently precise word? What exactly do I mean when I use the word? I suppose I mean 'good use'. But then what does 'good' mean? For me, this invokes notions of social justice, and finds more personal echoes in my political pedigree as a former Trade Union Shop Steward and Labour Party activist, as well as in aspects of my freelance media efforts. Certainly, politics and evaluation are inexorably entwined. Should some sense of 'social improvement', then, replace 'use' as the purpose of evaluation? Put another way, should House and Howe (1999) usurp Patton (1997) in my felt hierarchy of theorist construct import? If that were the case, where would my social work-derived 'service' commitment to 'stakeholders-as-clients' sit? Is it their sense of 'use' which matters most, or mine? After all, evaluators have an ethical responsibility to remain faithful to contractual obligations (see Chapter 5). And in determining these obligations, in undertaking our initial evaluability assessment, are we not 'free' to refuse the contract if we cannot agree a shared sense of purpose, if, say, the local programme particularities clash in some way with a broader sense of use as 'social improvement'? Then again, are the two necessarily mutually exclusive? In undertaking evaluations, are we limited in our range of learning or audiences? Could active public

disclosure through the media, including the internet, in addition to the more private disclosure of evaluation reports and academic publication, assume more prominence for evaluators, and render evaluation a broader social relevance? Has my thinking and questioning now come full circle to my construct starting point of delineating local instrumental use and broader social policy enlightenment use?

My intention with the last two paragraphs has not been to find definitional answers to the questions posed. They merely serve to illustrate the point that the search for philosophical meaning and practice consequence is boundless. Constructs are not permanent intellectual residences, but merely anchor points in a never-ending journey. What I understand evaluation to be now, may not be what I understand evaluation to be, say, three years from now. This, though, does not produce practice paralysis as Scriven seems to imply in complaining about the 'philosophically besotted' (Scriven, 1997, p.479). Rather, I am simply saying that philosophy matters:

...because it has important effects on how we envision and do our work. (Greene, 1996, p.537)

Chapter 3

INTEGRA – THE PROGRAMME

The world can only be grasped by action, not by contemplation....The hand is the cutting edge of the mind.

Jacob Bronowski, *The Ascent of Man*, 1973

INTRODUCTION

Based on the premise that offenders who find employment are less likely to re-offend than those who remain unemployed, the INTEGRA Project aimed to provide a comprehensive service to offenders in County Durham Probation Service (CDPS) whose access to employment is additionally hindered by their alcohol and/or other drug use. Although working in partnership with local voluntary agencies, the INTEGRA Project was a Criminal Justice project which supervised offenders as a result of community sentences or following release from prison. It was funded as a 2 year pilot by the European Social Fund under the INTEGRA strand of the EMPLOYMENT initiative.

This Chapter reviews research and policy in relation to offender employment, highlighting the Probation Service response, and noting European Union employment strategies and responses to the notion of social exclusion. The Chapter also considers research and theory relating to the 'link' between drug use and crime. This, then, establishes the research and policy backdrop to the INTEGRA Project.

The Project is then described in detail. How it is driven by employment/training outputs and underpinned by 'What Works' principles and values introduces key themes of the evaluation core design which will be considered further in Chapter 4. The Chapter concludes by constructing a programme 'logic model' for the INTEGRA Project at inception.

THE LINK BETWEEN EMPLOYMENT AND CRIME

The past two decades have seen increasing research interest into the impact of employment on recidivism, and this has been well documented by Downes (1993). Perhaps the authority most frequently quoted in the UK in recent years has been Lipsey's (1995) conclusion that:

The single most effective factor in reducing re-offending rates, with a positive effect size of 37 per cent, is employment. (p. 157)

This would certainly seem to validate anecdotal evidence of a strongly held belief among practicing Probation Officers that when offenders gain employment they are taking a significant step towards a crime-free life. However, while the large 'positive effect size' recorded by Lipsey is worth noting, the limitations of his wide-ranging meta-analysis should be recognised: Lipsey studied juveniles not adults, and in the USA not the UK. It is also worth noting his finding that employment interventions that did not lead to employment outcomes had a marginally detrimental rather than beneficial effect.

Perhaps the assessment which best provides a context for the INTEGRA Project comes from Farrington *et al* (1986). Their data were taken from the Cambridge Study in Delinquent Development, a long-term follow-up of 411 London males first studied by Professor D.J. West in the 1960s. The self-reported job history of each of the participants was analysed alongside their official criminal record, allowing a study of both convicted and unconvicted young men, and the effect on each of being in or out of work. This showed that the rate of offending during periods of unemployment was three times as great as during employment. Furthermore, the researchers were able to conclude that unemployment was related to crime independently of the many individual differences between convicted and unconvicted persons.

More detailed analysis also revealed that the type of offence was a key consideration. Crimes for material gain were associated with unemployment, but there was no such association for other crimes. Interestingly, too, they found that the benefit of being in work was reduced when the job was of a low status. And particularly significant from a Probation Service perspective was the finding that the effect of unemployment was strongest for those young men most predisposed to offending. Put another way, gaining meaningful

employment appears to considerably reduce a young male offender's likelihood of re-offending.

Of course, the significance of unemployment for older males and for female offenders is not embraced by the work of Farrington *et al*. Nevertheless, most agencies working to increase offender employment now infer a significance for older men, and adopt an equal opportunities stance in providing women with fair access to services.

THE PROBATION SERVICE RESPONSE

Despite the research of Farrington *et al* in 1986, and others (Box, 1987; Taring, 1982; Graham and Bowling, 1996; Farrall, 1995), it is, perhaps, surprising to note that many Probation Officers are still apparently reluctant to engage in employment work with offenders under their supervision. Bridges' (1998) thoughtful inquiry provides some insight into why this may be the case.

Until at least the mid 1980s, the non-engagement by most Probation Officers with employment work could be seen as a partial reflection of a lack of policy and direction at a national level. Other agencies, notably the Apex Trust and the National Association for the Care and Resettlement of Offenders (NACRO), were seen as the key players, providing a range of services, projects, and work schemes.

As the 1980s became the 1990s, the profile of employment work undertaken by the Probation Service nationally began to change. The Association of Chief Officers of Probation (ACOP) organised a national conference in July 1989, and the Home Office issued a letter to Chief Probation Officers in December 1990, recommending that they establish local offender employment forums. Shortly thereafter, a National Offender Employment Forum was also established, with regular attendance from a Home Office representative.

Throughout the early/mid 1990s, further advice and guidance on offender employment issues were provided by the then Employment Department, the Employment Service, Training and Enterprise Councils, and the Apex Trust. ACOP assisted in the drafting of much of this, and the emerging message was one of partnership working.

In May 1994, the Home Office issued a Probation Circular (40/1994) requiring that each Service establish a policy and strategy for its work in the areas of employment, training, and education of offenders. Working in partnership with others was again a key part of the message.

From 1993 to 1996, the Apex Trust was funded by the Home Office to provide consultancy to Probation Services in England and Wales, and to provide practical help, particularly in developing their partnership approaches with other agencies. In February 1995, the then Home Office Minister of State, Baroness Blatch, opened a multi-agency conference to promote partnership approaches.

During the final period of its consultancy, the Apex Trust was commissioned to produce a report assessing the progress made by the Probation Service and partner agencies in implementing the advice, guidance, and requirements of the previous few years (Apex Trust, 1996). The report's recommendation that:

...Probation services are the most appropriate agencies for taking the lead on offender employment issues...(p. 7)

was accepted by the National Offender Employment Forum as uncontroversial. In ten years, the Probation Service had moved from the marginal to the pivotal player in the offender employment world.

Nevertheless, Bridges' (1998) study of the case-files of over 700 offenders drawn from eleven Probation Service areas found that more than half the unemployed offenders failed to receive any intervention aimed specifically at improving their employability. As Bridges concludes:

There is substantial scope for undertaking employment interventions with many unemployed offenders who do not at present receive such help from the Probation Service. If this assistance were to be given, there is every prospect that a higher rate of success in terms of job starts and other employment outcomes could be achieved. There is good evidence – particularly from the work of Farrington *et al* and Lipsey - to suggest that this would lower the rate of offending of offenders subject to probation supervision. (p. 28)

THE EUROPEAN UNION RESPONSE

Since publication of the White Paper on 'Growth, Competitiveness, and Employment' in December 1993, the European Union has established a broad framework for a concerted attack on unemployment. The European employment strategy, agreed by the European Council in December 1994 in Essen, aims to make growth more employment-intensive, and includes the objective of facilitating the reintegration of those *socially excluded* groups at highest risk of long-term unemployment.

As Lister (2002) notes:

Social exclusion has become an increasingly fashionable concept in many parts of Europe, including New Labour's Britain. (p. 37)

However, it is worth pointing out that the concept of 'social exclusion' is contested in academic circles (Byrne, 1999; Levitas, 1998; Lister, 1996, 2002). Whilst the details of such debate are not necessary for this thesis, it is worth noting the broad thrust.

The notion of 'social exclusion', it is argued, implies a generally inclusive and cohesive society, and this is at odds with the realities of class and other divisions, as well as positing an over-simplistic polarisation between the 'included' and 'excluded'. In Europe and the UK, whilst exclusion is understood as multi-dimensional and embracing such 'problems' as unemployment, poor skills, lack of qualifications, low incomes, family breakdown, poor health, and poor housing, the focus for intervention policy is on employment. Paid work is seen as the way in which individuals can gain the resources, both material and social/cultural, to enhance their lives. Opponents, however, argue that, whilst not denying the relevance of such a 'distributional' material perspective, the intervention focus should shift to embrace in addition the 'relational' issues of participation and empowerment. In this way, the so-called 'excluded' could be drawn into the decision-making processes which affect their lives regardless of their poverty/work/welfare status.

With the INTEGRA Project in mind, it is important to remember that it was underpinned by the notion that paid work, or enhanced job prospects through skills attainment, would increase the 'social inclusiveness' of its beneficiaries. Essentially, it was an employment project and its outputs were thus likewise determined.

The European Social Fund (ESF) is the main financial tool through which the European Union translates its strategic employment policy aims into action. In the six years between 1994 and 1999, the ESF, which operates in all 15 Member States, spent ECU 47 billion (ESF website, accessed 4/2/99), accounting for almost ten per cent of the European Union's total budget. The ESF also helps to unlock funding at national level by means of a joint-funding principle which permits ESF support to be made available only for active measures already being undertaken by Member States to increase people's employment prospects. Essentially, then, the ESF provides funding for initiatives which develop or regenerate people's 'employability' and thereby promote their 'social inclusion'.

EMPLOYMENT was one such ESF funded initiative. EMPLOYMENT had four strands, each focusing on the reintegration of specific groups of disadvantaged people into the labour market:

- NOW** promoted equal opportunities;
- YOUTHSTART** targeted young people under 20;
- HORIZON** targeted disabled people;
- INTEGRA** targeted vulnerable/disadvantaged groups.

It was under the INTEGRA strand of the EMPLOYMENT initiative that the INTEGRA Project was funded for two years, funding which amounted to around £2.5 million. In providing such funding, the ESF recognised that the risk of permanent exclusion from the labour market is greater for those with the least education and training, the least appropriate skills or work experience, and the most frequent and lengthy periods of unemployment. Add to this the stigma arising from a criminal record, and/or that associated with problematic drug or alcohol use if disclosed, and the employment prospects for many Probation Service clients is poor.

Whilst the thrust of the INTEGRA Project was employment, it also held potential benefit for the Probation Service in achieving its goal of reducing re-offending by helping people overcome their drug difficulties, irrespective of whether they secured employment or gained qualifications/skills. Is this not the case?

That there is a link between some forms of drug use and crime seems obvious. However, the nature of this link, and the consequent focus for relevant agency intervention, is not so.

THE LINK BETWEEN DRUGS AND CRIME

At the outset of this brief section, I should point out two things: firstly, the research overview undertaken here owes much to the extensive work of Drugscope, the leading UK Charity in the field of drugs information; secondly, I am focusing on the links between drug use and *property crime*. To do so is not to deny that some drugs, especially alcohol, may facilitate violence – or, indeed, that some may even inhibit it - (Anglin and Speckart, 1988; Dobinson and Ward, 1986; Jarvis and Parker, 1989), or to deny the systemic violence associated with some forms of drug distribution (Goldstein, 1985). However, the hub of UK debate and concern, and much Criminal Justice activity, including the Probation Service, is the impact of drug use on such crimes as burglary, shoplifting, robbery and other theft.

As will be shown a little later in this section, research suggests that there is a link between drug use and property crime. But it appears that this link may arise in several ways (Coid *et al*, 2000; Best *et al*, 2001; Walters, 1998.):

- Drug use may lead directly to crime, say as a means of providing finance or as a result of the disinhibiting effect of some drugs.
- Crime may lead to drug use, say by providing disposable 'income' and contacts, or for coping with the stress of a chaotic criminal lifestyle.
- The interaction may be more complex, say whereby crime facilitates drug use, and then drug use prompts other forms of crime.
- There may be no causal links, but rather an association arising from a shared set of common causes.

Each of these explanations will apply to a greater or lesser extent for different people. What does the research suggest of the overall picture?

1. Around 4 million people use illicit drugs each year. Most illicit drug use is relatively controlled recreational use of cannabis and ecstasy. Whilst people who use illicit drugs are more likely to commit other forms of crime, there is no persuasive evidence of any causal link between drug use and property crime for the vast majority of this group (Ramsey *et al*, 2001; Flood-Page *et al*, 2000; Parker *et al*, 1998; Measham *et al*, 2001).

2. A very small proportion of all users – around 200,000 – have chaotic lifestyles involving dependent use of heroin, crack/cocaine, and other drugs. Of these, around 100,000 finance their use through crime. However, the majority of those who steal to buy drugs were involved in crime before their drug use became a problem (Bennett, 1998, 2000, 2001; Maden *et al*, 1991; Singleton *et al*, 1999; Lader *et al*, 2000; Hammersley *et al*, 1989; Edmunds *et al*, 1998, 1999; Gossop *et al*, 1998; Coid *et al*, 2000; Best *et al*, 2001)

3. If appropriate drug treatment is given to this group, they reduce their offending levels in the short term. (Gossop *et al*, 1998; Coid *et al*, 2000; Edmunds *et al*, 1998, 1999; Hearnden and Harocopos, 1999; Turnbull *et al*, 2000).

Chaotic drug users and persistent offenders, in contrast to controlled users and occasional petty offenders, appear to have limited social and economic resources, and limited exposure to legitimate life opportunities. The majority appear to emanate from deprived backgrounds, with inconsistent parenting, poor access to housing and health care, low educational attainment, and limited employment prospects (Harrison, 1992; MacGregor, 2000). Chaotic drug users, then, share the same constellation of risk factors that predict persistent involvement in crime, and exposure to 'social exclusion'.

Whilst such risk factors may predispose people to both chaotic drug use and to involvement in persistent offending, it is important also to realise how, once established, the two behaviours can become mutually sustaining (Walters, 1998; De Li Periu and Mackenzie, 2000). In this way, drug dependence tends to amplify and maintain the offending rates of people whose circumstances may predispose them to becoming persistent offenders. This has important policy and intervention implications. Whilst removing drugs from the equation may well result in reductions in offending in the short term, effectively maintaining any lifestyle changes as a result of treatment will also require holistic attempts to deal with the factors predisposing people to persistent offending, irrespective of any drug use.

INTEGRA – THE PROJECT DESCRIPTION

As mentioned in the Introduction to this Chapter, and based on the premise that offenders who find employment are less likely to re-offend than those who remain unemployed, the INTEGRA Project

aimed to provide a comprehensive service to offenders in County Durham Probation Service (CDPS) whose access to employment is additionally hindered by their alcohol and/or other drug use.

'WHAT WORKS'

Pressures on the Probation Service both to improve effectiveness and to demonstrate effectiveness have intensified over recent years, as outlined by Chapman and Hough (1998). According to these authors, a number of features in policy development have created these pressures:

From the beginning of the century until the early 1970s, the aims of the Service were undoubtedly rehabilitative, and work was focused on less serious offenders. The first major upheaval arose from the 'Nothing Works' disenchantment with the criminal justice system generally in the mid 1970s. However, this occurred at a time of little pressure on resources, and a government prepared to allow a high degree of local autonomy.

The second upheaval can also be traced to "Nothing Works" pessimism. If prison had no greater deterrent or rehabilitative impact than other penalties, there was an obvious cost case for rationing its use. Throughout the 1980s, government policy placed growing emphasis on diversion from custody, rekindling Home Office interest in, and expectations of, the Probation Service. Toward the end of the 1980s, the mission of the Service became the provision of alternatives to custody. The new demands on the Service were largely funded through budgetary expansion. During the 1980s, Probation growth outstripped all other criminal justice agencies, and demands for 'efficiency gains' were generally unheard of. But they were just around the corner.

The third upheaval came with the downward pressure on public expenditure in the early 1990s. Probation's period of budgetary expansion came to an abrupt end with the introduction of cash limits. At the same time, over the last decade, Probation's caseload has risen by almost 40%.

And now these pressures have coalesced with a reverse in the 'Nothing Works' pendulum. It is now the view, partly research-led, that Probation can do some things under some circumstances to change some offenders' behaviour. To protect the public, and to gain their support and confidence, Probation now has to (Chapman and Hough, 1998, p. 7):

- Ensure that existing knowledge about effective practice is consistently applied;
- Extend this knowledge base;
- Develop management systems for maximising effectiveness.

It is against such a 'What Works' backdrop that the INTEGRA Project was established, and this evaluation undertaken.

OPERATIONAL PRINCIPLES

The Project followed the 'What Works' principles put forward in Probation Circular 35/1998, *Effective Practice Initiative: A National Implementation Plan for the Effective Supervision of Offenders*. Identifying the key project components in relation to the nine principles of effective supervision of offenders:

- **Explicit empirically based model of change.** The Project espoused the motivational model of change, initially proposed by Prochaska and DiClemente (1984) – see below.
- **Targeting criminogenic needs.** The Project offered individually tailored programmes which included skills training, improving relationships, and stress management, all of which have been rated as having high levels of effectiveness in treating alcohol problems (Hodgson, 1994; Holder *et al*, 1991).
- **Responsivity principle.** Careful initial assessment aimed to ensure that the interventions used would be suited to the person affected. The developmental model cited above specifies relevant stages of change which also informed the assessment and intervention.
- **Effective methods.** Staff were trained in the required skills, and experience gained through the Project was integrated into their practice through a process of reflection and evaluation.
- **Skills oriented.** The Project's primary aim was to help people gain skills, and through these skills to gain employment.
- **Range of targets.** The criminogenic factors were tackled holistically.
- **Dose.** Because INTEGRA was one programme among all the services provided through CDPS, part of the evaluation was to explore its impact on its clients' use of other services, as well as the variation in the amounts of INTEGRA resources used by each client.

- **Ongoing monitoring.** This was supplied by the Project's information systems.
- **Ongoing evaluation.** Contracted to Centre for Applied Social Studies at Durham University.

Motivational Model of Change

Motivational approaches with problematic alcohol and other drug users were developed by Prochaska and DiClemente (1984). These place responsibility on the user to decide for himself or herself just how much of a problem there is, and then to choose what needs to be done about it. Their clinical research suggested that any resultant changes, internally attributed by the user in this way, tend to be longer lasting. As applied in the Project, the Project worker acts as a resource in the process, providing information and perspectives, alternatives and possibilities. This is not a passive role.

A key concept embedded within motivational approaches is that of cognitive dissonance. In essence, if a person perceives their behaviour to be at odds with their beliefs, attitudes, or feelings, then they experience discomfort – cognitive dissonance. A motivational condition is thereby created to bring about some change so that consistency can be restored. Within such a framework, the worker aims to increase the cognitive dissonance experienced by the user, and then help to direct this towards behaviour change, rather than:

- **MODIFIED BELIEFS (DENIAL)** – “It aint so bad.”
- **LOWERED SELF ESTEEM** – “I’m not worth it.”
- **LOWERED SELF EFFICACY** – “I can’t do it.”

Generally, the process of change can be conceptualised as embracing five stages:

- **PRECONTEMPLATION** – users do not perceive that they have a problem.
- **CONTEMPLATION** – they begin to weigh up the pros and cons of their alcohol or other drug use.
- **DECISION** – the point where users decide that, on balance, change is necessary.
- **ACTION** – the process of taking steps to put the change decision into effect.
- **MAINTENANCE** – involves the worker introducing and encouraging conscious coping strategies to help maintain the changes made, using Relapse Prevention techniques such

as those described by Marlatt and Gordon (1985). Occasional drink or drug-taking at this point does not necessarily constitute a return to previous behaviour, but should be understood as a lapse rather than relapse, and used for further learning. However, even if relapse does occur, it should still not be seen as a failure. Each revolution through this model of change should be used as a positive learning experience, hopefully leading to a greater chance of sustained behaviour change the next time around.

Moving successfully from one stage to another relies on a number of strategies:

- **CONSCIOUSNESS RAISING** – includes both feedback of specific information pertaining to the user, and education of the general effects of their behaviour.
- **SELF RE-EVALUATION** – includes an emotional and rational appraisal of both the pros and cons of trying to overcome problematic behaviour.
- **SOCIAL RE-EVALUATION** – focuses primarily on re-appraising the impact a problem has on others. Here, the values that might be essential to the user's sense of family or community are perceived as being in conflict with their behaviour.
- **SELF-LIBERATION** – involves changes at an experiential level that increase the user's ability to choose. Self-liberation involves them becoming aware of new alternatives for living, recognising any skills they possess to make alternatives achievable, or, indeed, the skills they may need to acquire and how to go about doing this.
- **SOCIAL LIBERATION** – involves changes in the environment which lead to more alternatives being made available.
- **COUNTER-CONDITIONING** – at the other end of the spectrum from 'changing through choosing' is a process where change is achieved by altering the conditional stimuli that exert some control over users' responses. When problematic responses are conditioned to such stimuli, then being conscious of the stimuli is not on its own likely to produce change, nor is conditioning likely to be overcome merely through the process of choosing to change. 'Gritting your teeth' is a notoriously poor method for most people. Instead, users can either alter the way they experience particular stimuli, or they can alter their environment to

minimise the likelihood of the stimuli occurring (stimulus control).

- **CONTINGENCY MANAGEMENT** – much of our behaviour is under the influence of its resulting consequences. If certain behaviours are rewarded in a meaningful way, then we are more likely to behave that way.
- **DRAMATIC RELIEF** – strong emotional reactions to events occurring in the environment can move people to change their ways, for example drink driving accidents and drug-related harm affecting someone close.
- **HELPING RELATIONSHIPS** – these can be both a pre-condition for change and a process of change.

Whilst making a previously confusing and demoralising insight to human motivation more understandable and predictable, this scheme also provides important practice guidelines. For instance, rushing to discuss the nuts and bolts of a detox regime with someone still in the **CONTEMPLATION** stage may not prove very productive. The aim, then, is to assess the appropriate focus and make the approach dependent on the stage the user is at. Many workers are experienced in helping people who have moved to the stage of **ACTION**, but there is also a real challenge in meeting the needs of the 'unmotivated', people who are at the pre-**ACTION** stages.

This last point raises interesting questions regarding *coercion* for any agency offering treatment within, or in partnership with, the Criminal Justice System.

A Few Words About Coercion and the Criminal Justice System

The Criminal Justice System, as Summers (2002 p. 224) notes, "essentially offers a coercive approach". Unlike drug users in the community who, it would seem, have the choice of seeking help or continuing with their current behaviour, the individual in contact with the Criminal Justice System does not usually have this option. The drug-using offender, for instance, may agree to volunteer for a community based Drug Treatment and Testing Order, but their choice is not 'treatment vs no treatment', it is 'treatment vs prison'.

On the face of it, this apparent Hobson's choice may seem to be at odds with an underlying tenet of the motivational model of change outlined in the previous section – that being to place responsibility on the user to decide for himself or herself just how much of a problem there is, and then to choose what needs to be done about it. Certainly, my own practice experience suggests that some

practitioners would be uncomfortable with any approach that involved coercion. Indeed, with INTEGRA in mind, and as will be discussed in later chapters, offering treatment within the control and enforcement imperative of the Probation Service resulted in some discomfort for Project staff, and particularly, it would seem, those seconded from partner voluntary agencies. However, as Marlowe *et al* (1996) point out, some form of coercion is commonplace at the point of entry to drug treatment services, and this is not always linked to legal mandate, but can be perceived as coming from other sources, such as family, friends, doctors, and employers. The issue, in my view, is not necessarily how people enter treatment, but what happens with them once they are there, and what sanctions may be imposed if they choose to leave.

Hough's (1996) review of the drug treatment literature suggests that the Criminal Justice System *can* effectively coerce people into treatment and keep them there. His review also suggests that the treatment outcome from such coerced entry is no less effective than that from 'voluntary' entry. However, Hough does raise interesting ethical questions about the nature of the treatment offered. Coerced entry into inappropriate treatment, for instance, may be argued as a miscarriage of justice and, today, in my opinion, could offer some grounds for action under the Human Rights Act (1998).

On the other hand, Hser *et al's* (1998) study indicated that individuals not complying with treatment referral were generally more dysfunctional, with higher levels of drug and alcohol use, and more severe degrees of psychological and family problems compared to those who did comply. For these people legal involvement was the significant factor related to any treatment entry. We should be mindful that those in greatest need may be the least likely to take up any form of treatment.

As Summers (2002) concludes,

The criminal justice system may offer opportunities to provide much needed help and this should be viewed optimistically and positively. But practitioners should also question where initiatives evolve from and be aware that there may be several agendas at play. The meaning of this in relation to practice issues, partnership, the focus of interventions, and the effect on the individual should continue to be a major focus for debate. (p. 231)

INTEGRA OPERATIONAL OUTLINE

The initial operational design of the Project embraced three phases:

- **the assessment phase** was designed to be delivered by the Assessment Team of County Durham Probation Service (CDPS). This included the preparation of pre-sentence reports on offenders, resettlement plans, supervision plans, and a specialist needs assessment, the Level of Service Inventory – revised (LSI-r) which highlights employment and drug use issues (Andrews and Bonta, 1995). Referral to INTEGRA for assessment was not a matter of choice for offenders, but was automatic if their LSI-r indicated that they were both unemployed and experiencing difficulties in relation to alcohol/drugs. There was no sanction if they then chose not to continue, but INTEGRA appointments were regarded as 'reporting contacts' under National Standards (Home Office, 2000) and failure to attend without good reason was noted as 'unexplained absence'. Two such occurrences usually resulted in a return to court for Breach proceedings, and the likelihood of further legal sanction, including the possibility of imprisonment.
- **the mentoring phase** of the Project was delivered by the specialist workers of the 'E Team'. This comprised a Manager, Senior Project Worker, three Probation workers, two seconded staff from partner agencies, an Information Assistant, and two administration workers. Based in Newton Aycliffe, the Team operated throughout County Durham and Darlington, working from CDPS premises, other community offices, and very occasionally visiting offenders in their homes. Staff completed a more detailed assessment report, as well as an action plan which detailed steps to be taken to allow offenders to sort out their lives and proceed to the third phase of the Project. Where appropriate, this phase of the Project could also involve referral to other specialist agencies.
- **the employment and training phase** was delivered by the staff of Astep, a project with a long record of providing services to unemployed offenders in the community. Astep provided guidance and support intended to ensure offenders were 'job-ready'. This could include assistance with job-seeking and application, as well as provision of, and help with access to, training opportunities and basic skill development.

Astep operated under the auspices of DISC, one of the partners to the Project. The other partners of the Project were the North East Council on Addictions (NECA) who provide specialist services in assessment, counselling and ongoing support, and NORCARE, a voluntary organisation who provide supported housing and services for disadvantaged people in the Northern Region. NECA and NORCARE each seconded a Project worker to the E Team. In addition, the Durham and Darlington Drug Action Team was represented on the management committee of the partnership in recognition of its central role in co-ordinating services for alcohol and other drug users in the County.

Team and Service Development

The staff of the E Team was appointed in June 1998, although the NECA secondee was not in place until September of that year. New premises were equipped and a programme of staff induction was implemented. This included team development events, motivational interviewing training, the creation of a more detailed assessment process, alcohol and drug training, IT training, and CDPS induction training for new and seconded staff.

The development of the mentoring phase involved negotiation of working protocols internally with the assessment team, the community supervision and resettlement teams, and externally with local prisons, NECA and NORCARE on the use of community facilities, the local Social Services Department, and the Drug Action Team on the development of the County criminal justice strategy.

It was anticipated that working methods would be subject to continual review and refinement through a regular series of professional meetings.

It should be noted at this point that, whilst ESF projects generally enjoy a 6 month development phase, this INTEGRA Project had closer to 6 weeks; the Project received an 11th hour go-ahead when a Belgian initiative withdrew. So, a lot had to happen in a relatively short space of time, including the evaluation design and its associated contract process.

Organisational Milieu

In June 1998, as the Project got underway, CDPS was reorganised. Previously, the work of the Service in assessing and supervising offenders was delivered in a generic way from community locations. The reorganisation resulted in the creation of functional units

delivering assessment, community supervision, and resettlement services for offenders as specialised services.

The assessment team, who were intended to form the first phase of the Project, used the Level of Service Inventory-revised (LSI-r) as an assessment tool. The LSI-r is an internationally recognised indicator of re-offending risk which uses a combination of static and dynamic risk predictors (Raynor *et al*, 2000). Among the criminogenic factors it identifies are alcohol and other drug use, education, and employment. New administration procedures were established county-wide to ensure that the results of these LSI-r assessments were translated into appropriate referrals to the mentoring phase of the Project.

Partnership Development

Effective partnership work involves the development of close working practices with wider services and agencies. INTEGRA partnership development included the formation of a partnership management group, the development of service level agreements for the secondment of staff, and the development of working agreements for the appointment and joint management of seconded staff. At inception the Project aimed to foster organisational learning by developing contacts with other relevant INTEGRA Networks and other specialist services within the Durham and Darlington agency network.

Project Output Targets

In line with the Project's aim to enable offenders to access employment and training, 'outputs' were geared to improving skills and referring offenders into employment and training. The Project set itself targets of 450 action plans in its first year, and 550 in the second. From this group they planned to enable 150 people (15%) to enter employment, 25 to enter further education, and 65 to enter workplace training. In addition, they aimed to obtain 50 Basic Skills Qualifications, 50 National Vocational Qualifications (NVQs), and 300 Key Skills Qualifications. These targets were based on an assessment of Astep's work with probationers at the time of application submission. This assessment was undertaken by an Assistant Chief Probation Officer who had subsequently moved to another Probation Service when the Project got underway.

Other outputs were aimed at developing links with employers through conferences, and the creation of a website and CD-ROM. The Project also sought to maximise the skills of project workers and the budget allowed for staff training.

Transnational Activity

An important dimension of the Project, and a requirement of ESF funding, was its transnational partnership, the ARTEMIS network, with projects in Bolzano (Italy) and Santarem (Portugal). These provided contrasting cultural, political and economic contexts, with different criminal justice contexts, and different experiences of how national and local government, as well as voluntary organisations, combine to provide services.

The Italian Project worked with prisoners pre- and post-release to identify educational and training needs. Although the Portuguese partners worked with a wider catchment group, they also developed motivational programmes, working in a similar way to DISC, both factors that offered opportunities to compare and contrast practice.

At the first transnational conference hosted by the E Team in November 1998, the partners agreed to focus their shared learning on five areas of mutual interest:

- motivational interviewing,
- dissemination and links with local agencies,
- attitudes of employers,
- social exclusion, and
- working methods.

The partners agreed to share information, lessons learned, documentation, and instruments/questionnaires developed in these areas of work. Exchange visits of staff were organised to focus on these areas, providing structure and an additional method of dissemination. Further, the transnational conferences were seen to offer a mechanism to review progress and arrangements.

Summarising the Project and its operation, then...

...INTEGRA PROGRAMME LOGIC MODEL

Fig. 3.1 attempts to summarise the Project in the form of a 'programme logic model'. Put simply, a programme logic model is a systematic summary picture of how a programme works, or how it is intended to work:

The program logic model is defined as a picture of how your organization does its work – the theory and assumptions underlying the program. A program logic model links outcomes (both short- and long-term) with program activities/processes and the theoretical assumptions/principles of the program. (W.K. Kellogg Foundation, 2001, p. 7)

The term 'logic model' is frequently used interchangeably with the term 'programme theory' in the evaluation field. Logic models can alternatively be referred to as 'theory' because they describe how a programme works and to what end:

- A plausible and sensible model of how a program is supposed to work. (Bickman, 1987, p. 5).
- The set of assumptions about the relationships between the strategy and tactics the program has adopted and the social benefits it is expected to produce. (Rossi, Freeman, & Lipsey, 1999, p.98).
- The full chain of objectives that links inputs to activities, activities to outputs, outputs to outcomes, and outcomes to ultimate goals constitutes a program's theory (Patton, 1997, p. 218).
- A set of interrelated assumptions, principles, and/or propositions to explain or guide social actions. (Chen, 1990, p. 40).
- An explanation of the causal links that tie program inputs to expected program outputs. (Weiss, 1998, p. 55).
- A chain of causal assumptions linking program resources, activities, intermediate outcomes, and ultimate goals (Wholey, 1987, p. 78).

Whilst there is no best way to construct a logic model, they generally fall into one of three categories, or hybrids thereof:

- **The theory approach** - links theoretical ideas together to explain underlying programme assumptions. The focus here is on the problem or issue and the reasons for proposing the solution suggested in the programme's approach. The theory logic model is broad and about 'big ideas' not about specific programme 'nuts and bolts':
- **The outcome approach** - displays the interrelationships between specific programme activities and their outcomes.

- **The activities approach** – like the outcome approach, also connects programme resources and activities to desired results but does so in much more detail.

The model of the INTEGRA Project represented in Fig. 3.1 overpage is essentially a hybrid of the *conceptual* theory and the *applied* outcome approaches. It represents the *intended* Project at inception. Later in this thesis I will redraw the model post-evaluation as a way of summarising what the evaluation suggests *actually* happened.

INPUTS	ACTIVITIES	OUTPUTS	OUTCOMES
<p>£2.5 million; Staff Partners Accommodation Equipment Materials</p>	<p>Automatic referral to CDPS Assessment Team for all offenders whose LSI-r notes both unemployment and drink/drug issue</p> <p>Following this initial assessment, referral to E Team Mentoring Phase for fuller assessment if felt to be appropriate and offender willing.</p> <p>E Team agree action plan with offender to help them overcome drink/drug difficulties. This work to involve other specialist agencies if appropriate. Therapeutic work ceases as soon as an employment output is achieved. Offender referred to Astep phase for their specialist assistance in job finding and or training when felt to be ready.</p> <p>IN ADDITION Share knowledge with transnational partners; Link with other relevant INTEGRA Networks; Disseminate knowledge gained, incl. with employers; Train staff.</p>	<p>450 action plans Yr 1 550 action plans Yr 2 150 into employment; 25 into further education; 66 into vocational training; 50 Basic Skills Referrals; 50 NVQs; 300 Key Skill Quals.</p> <p>IN ADDITION Conferences; Website; CD-ROM; Staff Quals.</p>	<p>REDUCED REOFFENDING AMONG BENEFICIARIES.</p> <p>ENHANCED SKILL AND KNOWLEDGE IN LOCAL CRIMINAL JUSTICE AND DRUG TREATMENT NETWORKS.</p> <p>ENHANCED UNDERSTANDING AMONG LOCAL EMPLOYERS.</p>
<p>PREMISE</p>	<p>Offenders who find a job are less likely to reoffend; Overcoming drink/drug problems enhances employability and may reduce reoffending risk directly; Training/qualifications enhance employability.</p>	<p>CONTEXT</p>	<p>Probation Service; Drug Treatment Network; Employment Market;</p>

Fig. 3.1 INTEGRA Programme Logic Model

Chapter 4

THE TRUTH I AM TRYING TO GRASP...

...is the grasp that is trying to grasp it.

R.D. Laing, *The Politics of Experience*, 1967

INTRODUCTION

The previous chapter reviewed the research and policy backdrop to the INTEGRA Project, described the Project, and concluded by developing a hybrid programme logic model to summarise theory, process and outcomes. Chapter 2 explored the multi-faceted landscape of evaluation, and situated my personal stance.

This methodology chapter embraces a fusion of these two. As such, it seeks to delineate the design and then the process of this programme evaluation, to describe the planning and then the action buttressing its construction. However, as the corner-stone of this thesis, the function is also to consolidate the whole. This requires that I additionally sow some seeds for what is to follow in subsequent chapters. I will continue my 'personalising' (Kushner, 2000) approach, outlining aspects of my studentship role in contributing to a pre-existing 'core' methodology. Whilst there is some discussion throughout regarding the appropriateness of certain decisions taken in relation to design and process, this merely serves as a pointer to the more focused discussion provided by Chapter 8.

BACKDROP

As outlined in Chapter 2, the evaluation field is contentiously diverse, and as the goals and practices of the field have diversified, so too have the evaluators' roles and relationships with the programmes they evaluate. For instance, those operating within traditional tenets and holding a summative judgement view of evaluation will more than likely favour a distanced relationship, and

rely on methodological rigour, rather than a highly interactive and interpersonal approach.

Evaluators, then, can hold all manner of philosophical and methodological preferences and, even, prejudices. Choosing an *external* evaluator to match programme and budget can represent an intimidating choice for any conscientious and considered programme director or evaluation taskforce. One of the most basic role divisions for evaluators is, arguably, that between those who are 'academic orientated' and those who are 'service orientated'.

Before they merged into the American Evaluation Association (AEA) on January 1st 1986, a stratified random sample of the members of the Evaluation Network (ENet) and the Evaluation Research Society (ERS) were surveyed by Shadish and Epstein (1987). These were the two major evaluation organisations at the time of the survey. The total population included 1,864 members of the ERS, and 1,150 members of Enet, after eliminating evaluators who lived in foreign (i.e. outside the US) countries. A total of 604 respondents were randomly sampled, stratified by the size of the organisation, with duplicate names eliminated randomly from one or other list. 350 subjects (57%) responded to the first wave of postal questionnaires, and 53 (9%) to a second wave of mailings aimed at initial non-respondents. Shadish and Epstein enquired about a variety of issues related to evaluators' values and practices. Responses clustered around two contrasting views of evaluation: *academic evaluators* and *service evaluators*:

The general discrepancy between service-oriented and academically oriented evaluators seems warranted on both theoretical and empirical grounds. (Shadish and Epstein, 1987, p.587)

According to Shadish and Epstein, in the US at least, *academic evaluators* tend to be at universities and stress the research purposes of evaluation, employing traditional standards of methodological rigour, summative outcome studies, and contributions to social science theory; *service evaluators*, on the other hand, tend to be independent consultants or internal evaluators and stress serving stakeholder needs, assisting programme improvement and decision-making, and formative approaches.

In addition, they found that the 31% of the respondents who described their primary professional identity as that of 'evaluator'

were more likely to exhibit service evaluator tendencies. Conversely, those who took their primary professional identity from their academic disciplines were significantly more likely to be engaged in academic evaluative research, emphasising research outcomes and summative judgement (p. 581).

Although this is now 'old' research, and although it is North American, I believe it still holds relevance for evaluation as a discipline in the UK, and, more specifically, for this INTEGRA programme evaluation.

Based on my project management background, I would characterise myself as a *service* evaluator by inclination. The manager of the evaluation who was the originator, to a great extent, applied an *academic* approach, based on her previous experience of evaluation research.

Of course, the design and process of this INTEGRA evaluation depends on much more than the evaluators' philosophical and methodological preferences or prejudices. Nevertheless, it is a central dimension, and one which any programme considering the selection of an *external* evaluator should seriously consider. In the evaluation guide for their own programme directors, the W.K. Kellogg Foundation puts it this way...

Even more powerful than the notion that there are different paradigms with different assumptions about the world and how it works...is *how much our particular paradigms/assumptions influence the questions we ask; what we think is important to know; the evaluation methods we use; the data we collect; even the interpretations and conclusions we make.*
(W.K. Kellogg Foundation, 1998, p. 8)

THE STUDENTSHIP & MY ROLES

This evaluation, and its accompanying PhD studentship, has been a constant compromise between what was pragmatically and academically possible within the resources available, and within the constraints imposed by my power and status, and by my learning and development. Of particular significance have been my multiple roles.

I began this enterprise as a 'Research Assistant', working to an agreed evaluation design, a design, as I will outline later, that was

summative and experimental. Intuitively, though, I also brought an experienced project manager's *service* preference, and all what that implies: serving stakeholder needs, assisting programme improvement and decision-making, and preferring formative approaches.

Of course, serving stakeholder needs also means sticking by agreements already made. If a summative experimental design was what had been contracted, then that is what should be undertaken. While retaining a commitment to flexibility, any serious unravelling of the agreed design would be a substantial undertaking, not least given the time limits. The Project was already 6 months into operation when I started, and had a further 15 months to run.

At the same time, although very much a whisper for the first year, was a constant academic voice, a constant reminder that I had to use this experience, and bring sufficient of my own to it, to satisfy the requirements of a thesis – I was a 'student', with a lot to learn.

Then, a few months into the studentship, two things happened, almost in chorus, and two things that were to shift some of the fog from my brain and some of the shackles I felt myself to be towing.

The first involved a handshake from my evaluation manager, and the accompanying words: "Congratulations. You're now the 'lead researcher'". A shift in roles and I felt more empowered.

The second was a thorough reading of Patton's 'Utilization-focused Evaluation' (1997). Here was something I could use to develop a framework for this evaluation, a framework to understand my contribution to it, and a framework to examine the stance I was beginning to take in relation to my role as 'evaluator'.

Meaning and Metaphor

In his Chapter 2 section, 'Creative Beginnings' (pp. 33-38), Patton discusses the value of metaphor in helping us 'make connections between seemingly unconnected things, thereby opening up new possibilities by unveiling what had been undetected' (pp. 33-34). He also describes the work of Smith (1981) who directed a 'Research on Evaluation' Programme, and in which he and others thought about evaluators as such things as poets, architects, and photographers. 'They consciously and creatively used metaphors and analogies to understand and elaborate the many functions of program evaluation' (Patton 1997 p. 34).

All of which got me to thinking about the amoeba as a metaphor:

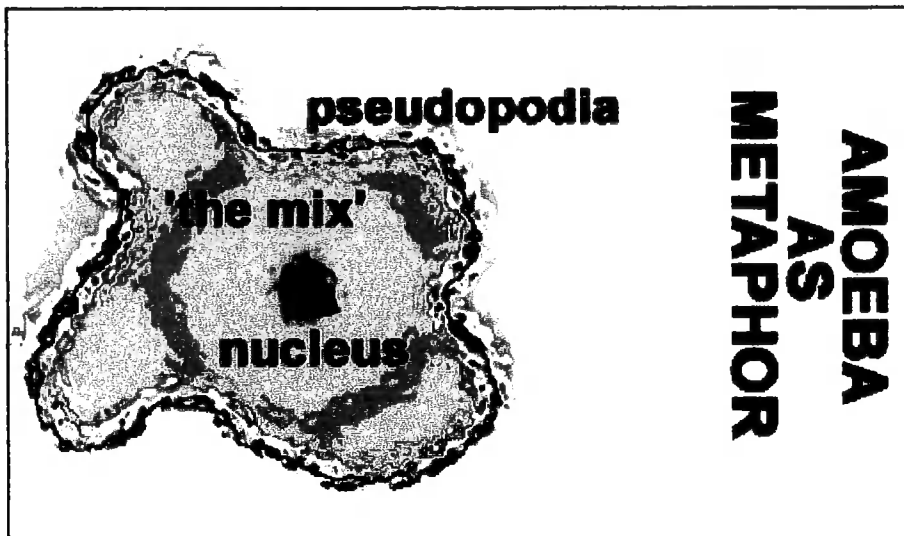


Fig. 4.1: Amoeba as Metaphor

amoeba (@'mi:b@). *Zool.* Pl. **amœbæ**, **amœbas**.

[ad. Gr. ἀμοιβή change, alternation.]

A microscopic animalcule (class *Protozoa*) consisting of a single cell of gelatinous sarcodæ, the outer layer of which is highly extensible and contractile, and the inner fluid and mobile, so that the shape of the animal is perpetually changing. (Oxford English Dictionary on CD ROM, version 1.13, Oxford University Press, 1994)

Simultaneously loose and tight, the amoeba (Fig. 4.1) offers a suitable metaphor to explore the INTEGRA evaluation. At core, we have a tight nucleus: the summative experimental core design and the Project/Research Management Group, both driving the motion. In there, too, I would put the ethical and practice principles which guide the development and delivery of the evaluation. Surrounding this, a mobile fluid mix, a nourishing concoction of support, sense-making, information-sharing, and data analysis. This would include such things as: my supervision and training, any books and people available to me, as well as the *Project and its staff*. And at the periphery, an everchanging membrane, a place where ideas and plans coming from within can be given pseudopodic and systematic evaluation shape, and where data, experiences, and possibilities from the *Project operation and its context* can permeate back into the mix and then feed the nucleus. The overall enterprise, then, is organic, it is active, reactive and adaptive, with a relatively stable

guiding nucleus, and the INTEGRA Project itself is at once both inside and outside the membrane, and, via its Project management, both inside and outside the nucleus. All of this implies a highly interactive relationship between the Project and the evaluation process, between the Project and myself.

Clearly, as outlined in Chapter 2, such an evaluative stance would raise objections from traditional evaluation theorists, like Campbell and Scriven, who teach that data collection, for the results to be valid and reliable, should be separate from the programme. They would also exhort that a close relationship between the evaluator and the programme could undermine the evaluator's neutrality and independence:

It's quite common for younger evaluators to 'go native', that is, psychologically join the staff of the program they are supposed to be evaluating and become advocates instead of evaluators. (Scriven, 1991, cited in Patton, 1997 p. 113)

Indeed, this was just such a concern expressed by my evaluation manager. The validity and relevance of this concern for this evaluation will be discussed further in Chapter 8.

For now, I should say that my approach throughout this evaluation relied on a number of guiding principles (Shadish *et al.* 1995):

- **Systematic Enquiry** – evaluators conduct systematic, data-based inquiries about what is being evaluated;
- **Competence** – evaluators provide competent performance to stakeholders;
- **Integrity/Honesty** – evaluators ensure the honesty and integrity of the entire evaluation process;
- **Respect for People** – evaluators respect the security, dignity, and self-worth of the respondents, programme participants, clients, and other stakeholders with whom they interact;
- **Responsibilities for General and Public Welfare** – evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

In general terms, of course, these principles may well seem quite modest, but render each in its negative and their influence should become clearer.

In more specific terms, my approach has been driven by Patton's (1997) 'utilization focus' – as outlined in Chapter 2. This too, would be the focus claimed by my evaluation manager, and not least in driving her approach to the core design of this evaluation – a core design I inherited on taking up my studentship 6 months after Project start.

EVALUATION DESIGN

One of the most challenging, and yet I would also argue exciting, aspects of programme evaluation is that there should be no 'right' way to do it, no 'off the shelf package' to pick up and apply in all situations. Being active-reactive-adaptive, being situationally-responsive, means that evaluators do not impose 'cook-book' designs. Ideally, each evaluation should be made to order. Evaluators:

...don't do the same thing time after time. They are genuinely immersed in the challenges of each new setting and authentically responsive to the intended users of each new evaluation. (Patton, 1997, p.135)

Such a bespoke approach implies a thorough planning process, where stakeholders and evaluator(s) come together to negotiate clarity of purpose and method. This is the hub around which evaluation revolves. In tailoring an evaluation this way, three particular factors bring weight to the design questions which need formulating, and which then need answers before proceeding.

THREE FACTORS INFLUENCING DESIGN QUESTIONS

1. Who Are The Stakeholders?

All evaluations can be seen to have multiple stakeholders, the people or groups who have a stake, a vested interest, in the evaluation or in its results. Apart from the obvious cast of: funders, programme administrators and staff, the user participants, and the evaluators themselves, the line-up could also include anyone with a more indirect interest in the programme's effectiveness, such as journalists and the public at large. Indeed, these last two bring a significant contribution the INTEGRA evaluation design – a point to which I will return later in this chapter.

Clearly, then, stakeholders can have diverse and even competing interests, as evident in the potential questions each might wish to see answered by any evaluation:

No evaluation can answer all potential questions equally well. This means that some process is necessary for narrowing the range of possible questions to focus the evaluation. In utilization-focused evaluation, this process begins by narrowing the list of potential stakeholders to a much shorter, more specific group of *primary intended users* [emphasis added]. Their information needs, that is, their *intended uses*, focus the evaluation. (Patton, 1997, p. 42)

As seen from the previous chapter, major stakeholders, but not necessarily Patton's *primary intended users*, in the INTEGRA evaluation include: County Durham Probation Service (CDPS), the European Social Fund (ESF), partner agencies, Project staff, and the evaluators. With the direct exception of ESF, each of these were members of the CDPS-chaired 'Research Management Group', the *evaluation taskforce* which oversaw all aspects of this work:

Where possible and practical, an evaluation task force can be organized to make major decisions about the focus, methods, and purpose of the evaluation. The task force is a vehicle for actively involving key stakeholders in the evaluation...The task force allows the evaluator to share responsibility for decision making by providing a forum for the political and practical perspectives that best come from those stakeholders who will ultimately be involved in using the evaluation. (Patton, 1997, pp. 352-353)

These too, with the exception of myself, were active participants, to a greater or lesser extent, in the 'core' design process.

Beyond the contributions of the stakeholders, including the evaluator(s) and their:

...personal knowledge, skills, style, values and ethics.
(Patton, 1997, p.136)

and beyond the availability of an adequate budget, what other influences might bear on the process of determining the purpose of the evaluation and the shaping of suitable evaluation questions?

2. The Programme Policy and Practice Context

Principally, this embraces what has already been outlined in Chapter 3, the policy and practice backdrop to the development and operation of the INTEGRA Project. Clearly, these issues also have a direct bearing on the questions brought by the stakeholders to this evaluation design. However, rather than lump them into the stakeholder paragraphs, I feel they deserve separate summary.

As Chapman and Hough (1998) describe, pressure is being increasingly applied to all Probation Service areas to conduct ongoing evaluations of the effectiveness of their interventions with offenders. Once successful interventions have been identified, once they have 'proven' to be effective, they will form part of a developing menu of effective programmes. It is not surprising that, faced with such an intense 'proving' backdrop, evidence-based approaches to Probation policy and practice development increasingly tend to adopt a methods focus derived from the positivist traditions (Nutley and Davies, 2000).

In the past, what 'rigorous' evaluations have been undertaken have relied largely on reconviction rates as a measure of their effectiveness i.e. as a proxy measure for re-offending (for example, Wilkinson and Morgan, 1995; Roshier, 1995; Oldfield, 1996; Richards, 1996). While reconviction rates are arguably the most relevant available 'measure' to address the issue of 'proving' Probation effectiveness, they also present a number of disadvantages (as reviewed by, amongst others, Mair and Nee, 1992; Lloyd *et al.*, 1994). For example:

- They are vulnerable to local differences in prosecution practice and clear-up rates;
- They need unpacking to explore other outcomes such as the seriousness of offence and resulting sentence;
- They often include 'pseudo-convictions', offences committed outside the time period under study, but with court dates inside;
- They need quite large numbers of offenders if comparisons are to be meaningful;
- Their use depends on knowing what reconviction rate would have been likely in the absence of the programme intervention. This calls for well controlled evaluation designs through the use of matched comparison groups or reliable baseline predictors of reconviction risk.



Given such difficulties, there is now widespread interest in other possible measures of impact, such as attitudes and social problems, (Raynor, 1998) which might provide more relevant proof of 'What Works' (McGuire, 1995), and help Probation Officers meet the requirements of Home Office National Standards (Home Office, 2000) to assess offenders' 'suitability' for particular forms of community sentence and supervision.

Nevertheless, and whatever the outcome measures used, the pressure remains within the Probation Service to prove effectiveness, and statistical results are the preferred standard for validity and meaning.

Of course, the Probation Service is not alone in the UK public sector in holding such a preference. This also dominates health care (Mays and Pope, 1996), and alcohol treatment services are no exception (Wright, 1999).

Given the INTEGRA Project's remit of targeting unemployed alcohol and other drug-using offenders, the influence of methodological preference from the alcohol arena could be expected. However, as outlined below, the illicit drugs field can be an entirely different matter.

3. The Programme Users' Context

Perhaps not surprisingly, problematic drug users in treatment tend to be available for 'research' approaches (Rosenblum *et al.*, 1995; Bloor *et al.*, 1997; Parker and Kirby, 1996). Being in treatment implies both contactability, perhaps 'inpatient' or via their attendance at a treatment centre, and, presumably but not always, some degree of motivation. The methodological problems arise for those drug users where the potential cost of exposure or scrutiny is high for them:

Those who live drugs - crime lifestyles are not just paranoid. People really are out to get them. Being routinely in possession of quantities of Class A drugs, supplying them and regularly committing acquisitive crime make such users justifiable policing targets. Then there is the legion of local officials to be dodged as they chase benefit fraud, rent arrears, council tax evasion and unpaid higher purchase payment.

Unsurprisingly social researchers from the university or social research company enquiring about drug use patterns are unlikely to receive an audience, making traditional survey results a poor indicator of drug use in the shadows and at the margins. (Brain *et al.*, 1998).

Just how far might E Team probationers see themselves as being in 'treatment'? As described in Chapter 3, referral to the Team was not a question of choice. Using the developmental model of change (Prochaska and DiClemente, 1984) within the body of the Project theory, one could anticipate a large number of 'precontemplators' among those referred. Precontemplators, as again noted in Chapter 3, are significantly less susceptible to change strategies than people at other stages of the change process: they grasp less information about their problems; they spend less time and energy re-evaluating themselves; they experience fewer emotional reactions to the negative aspects of their behaviour; they are less open with others; and they do little to shift their attention or their environment in the direction of overcoming their problems. Moreover, those within the Criminal Justice System may understandably feel coerced, and merely hopping through a hoop of pacifying powerful 'others' to take the pressure off themselves. All of which begs a question about their likely willingness meaningfully to participate in anything other than their statutory obligations.

Furthermore, precontemplators may well still be using drugs problematically, with all that implies, not only in what they may feel themselves to be risking by exposure within the Criminal Justice System, but also prone to problems of memory, the influence of binges, vagaries in supply and consequent withdrawal difficulties, as well as irregular and unpredictable 'income opportunities'. All of which raises another very pragmatic question about their capacity to remember, and be available for, research appointments. – not least appointments offered at a probation office, a point to which I will return later.

Clearly, accessing drug users in the Probation Service poses a significant methodological conundrum for a suitable evaluation design – not least, in the absence of a convenient and user-friendly 'gathering' point. (As outlined in Chapter 3, probationer contact with the INTEGRA Project occurred throughout County Durham on an appointment basis at probation offices.)

Paradoxically, in contrast to its Probation 'What Works and proving' heritage, and as Chapter 3 again outlines, the INTEGRA Project also inherited the potential for a 'learning and improving' agenda from its European Social Fund policy roots. Funded for two years, it was charged with being *innovative*, developing models of good practice which might impact on national policy and/or be incorporated into existing mainstream provision. The question of how much scope for innovation, within the timescale, existed for the Project, and a reframing of the notion of 'innovation' from 'the development of new practice' to 'the resolving of paradoxes and contradictions', will be discussed in Chapter 8. The question for now is how much weight might this bring to the evaluation design?

At core, and as detailed in the next section, the evaluation design can be best described as summative and experimental. This would fit the 'What Works' orientation, but not the ESF perspective for new time-limited programmes charged with *innovation*, not least as their programme design should remain open to change and adaptation in the light of learning:

Sometimes evaluations of new programs are expected to address questions of impact and efficiency, but the unsettled nature of the programs...makes those issues premature...[rather] when new programs are initiated, especially innovative ones, evaluation is often requested to examine the social needs the program should address, the program design and objectives, the definition of its target population, the expected outcomes, and the means by which it assumes those outcomes can be attained...formative evaluation is likely to be more apt in these cases. (Rossi, Freeman & Lipsey, 1999, pp. 44-45).

For instance, had the Project started out working one-to-one with offenders and then had taken a therapeutic decision to introduce groupwork for suitable candidates, the overall evaluation results could only indicate the effects of attending INTEGRA relative to any control. Because offenders would not be randomly assigned for groupwork, any other intervention comparisons would be contaminated by selection bias:

If the program has changed appreciably before outcomes are measured on the participants, the effects of the different variants of the intervention are all mixed together in the results, and there is no easy way to determine what effects are produced by any given form of the intervention. (Rossi, Freeman, and Lipsey, 1999 p. 301).

Interestingly, my first few meetings with the INTEGRA Project staff demonstrated that, whilst aware of the evaluation design, they had similar reservations about it too. This group of stakeholders was concerned that any lessons to be learned would come too late to be of any use to them. They were expected, after all, to be innovative. What they were looking for was something more formative. They were also concerned that some of the work they did, some of their performance, would get "lost between the instrument tick boxes". They wanted something more qualitative. And finally, but very pragmatically, they were concerned that the entire evaluation could go to waste by my not seeing sufficient numbers of their 'hard-to-reach' clients to satisfy the statistical demands of the experimental design.

PUTTING IT ALL TOGETHER

An evaluation design, then, is heavily contextualised by the nature of the programme and its intended beneficiaries, by its political and policy backdrop, by the interests of its stakeholders, and by the experience and preferences of the evaluator(s). All of which call for careful assessment of the landscape before design launch.

Such a careful reconnaissance, though, requires time, and that is one resource, as noted in Chapter 3, that this programme and its evaluation planning did not have. So, a lot had to happen in a relatively short space of time, including this evaluation design and its associated contract process.

The substance of the evaluation planning fell to Dr. Justine Schneider from the Centre for Applied Social Studies (CASS), and Senior Probation Officer Mike Creedon, the INTEGRA Project Manager. As part of their planning, some discussion took place, too, with other major stakeholders, under the auspices of the Probation-chaired Project Management Group, and particularly with members from the UK voluntary sector partners: DISC, NECA, and NORCARE. Fortunately for the evaluation planning, as outlined in Chapter 3, the Project also benefited from a clear sense of what it

was aiming to achieve, what theory drove those aims, and what 'measures' the ESF co-sponsor would require for monitoring purposes. Out of this rapid mix, then, came an agreed set of evaluation questions to underpin the design.

Evaluation Questions

In line with the Project's aim to help offenders access employment and training, outputs were geared to improving skills and to securing jobs and training. Consequently, fundamental (summative) evaluation questions were derived from these:

- How many service users enter education or training?
- How many complete education or training courses?
- How many people obtain jobs?
- How much time after INTEGRA is spent in employment?

Bearing in mind the Probation Service 'What Works' imperative, and particularly the advantages of those outcomes, as mentioned earlier, that do not suffer the same methodological difficulties as 'reconviction', other evaluation questions were derived from these:

What is the effect of the INTEGRA programme on service users':

- Attitudes to offending?
- Level of skills?
- Self-esteem?
- Attitudes to work?
- Alcohol and/or other drug use?

Additionally, the evaluation sought to answer:

- What are the mean costs of given outcomes for these service users?
- What factors predict successful outputs?
- For successful users, what factors predict higher costs?
- How satisfied are staff with the INTEGRA Project and their role in it?
- How satisfied are service users with the Project?

To answer these questions, a summative, longitudinal, and experimental⁸ design was proposed and accepted, and a number of standardised questionnaire measures were chosen. These are summarised in Figure 4.2, and described below.

Fig. 4.2: Forms and Functions – Service Users

Outcome Domains	Questionnaires for Administration to Service Users
Risk of re-offending and needs.	Level of Service Inventory – revised, [LSI-r] (Andrews and Bonta, 1995)
Use of services.	Client Service and Socio-demographic Receipt Inventory, [CSSRI] (Beecham, 1995)
Offenders' work attitudes, anxiety and competence.	Work Behaviours and Attitudes Scale, [WBAS] (Harris <i>et al.</i> , 1993)
Offenders' self-esteem and perceptions of 'empowerment'.	Empowerment Scale , (Rogers <i>et al.</i> , 1997)
Offenders' drug use, health, and social functioning.	Maudsley Addiction Profile, [MAP] (Marsden <i>et al.</i> , 1998)
Offenders' therapeutic relationship with Project staff.	Working Alliance Inventory, [WAI] (Horvath and Greenberg, 1989)
Offenders' subjective experience of the Project.	Balance Sheet (in users' words) of pros and cons of INTEGRA contact.

The Level of Service Inventory – revised (LSI-r) is completed for all probationers in County Durham, and the results for those offenders not referred to INTEGRA form the basis of the evaluation's 'control' data. The LSI-r assesses 'risk' and 'needs', and rates a person's criminal history, education, employment, financial situation, family relationships, accommodation, leisure time use, companions, alcohol/drug problem, emotional state, and attitudes towards crime and Probation. Across these domains, the LSI-r generates a summary score which can be used as an indicator of change. There is a growing body of evidence indicating

⁸ The appropriate characterisation of the INTEGRA evaluation core design (a core design developed by the Evaluation Manager and which I inherited on my appointment) as either 'experimental' or 'quasi-experimental' (or even 'pre-experimental,') has been the subject of some debate, and, for me, the source of some confusion. Appendix I seeks to clarify the key issues.

the validity of the LSI-r in predicting risk. The LSI-r manual (Andrews and Bonta, 1995) cites over 20 studies in North America covering thousands of offenders during the 1980s and early 1990s which demonstrate relationships between LSI assessments and a number of correctionally relevant outcomes including reconviction, self-reported offending, parole outcome, breaches of prison rules and violations of supervision requirements. In addition, an independent meta-analysis of recidivism predictors in use in North America found the LSI-r to be the most accurate of those studied (Gendreau *et al.*, 1995). A more recent UK Home Office study (Raynor *et al.*, 2000) examines over 2,000 offenders who were assessed using the LSI-r or another assessment instrument, ACE - Assessment, Case management and Evaluation. The focus was to examine the instruments' ability to predict risk of reconviction, to reliably assess risk and needs of the offender and to measure any changes in these factors over time. The study finds that both assessment instruments are able to predict reconviction at a much higher than chance level and have good reliability (the reader is directed to the source for reliability coefficients). However, the LSI-r, which has a longer history of development, provides a slightly more accurate risk assessment than ACE⁹. As outlined in Chapter 3, any probationer identified by the LSI-r as being both unemployed and having an alcohol and/or other drug problem was automatically referred to INTEGRA.

The Client Service and Sociodemographic Receipt Inventory (CSSRI) was developed at the Personal Social Services Research Unit (PSSRU) at the University of Kent. It gathers information about all services received by a service user over the past three months. When costed, this provides a measure for the package of inputs received, from all sources, and by every respondent. The CSSRI was developed from earlier PSSRU research on child care and young offender services in the early 1980s, and on the Economic Questionnaire of Weisbrod *et al.* (1980). The CSSRI is now widely used in cost research, particularly in the health field.

As Merrington and Hine (2001) note, most Probation evaluations do not currently consider whether a programme represents good value for money. However, the picture is changing. The Crime Reduction Programme Guidance Note 1 (Dhiri and Brand, 1999),

⁹ At the time of thesis submission, Durham Probation staff, along with colleagues in Kent, Hertfordshire, Cambridgeshire and Bedfordshire, are piloting the use of a new national system for assessing the risks and needs of an offender. The Offender Assessment System (OASys) has been jointly designed by the Prison and Probation Services.

provides a useful introduction to costing within the Criminal Justice System and a cost template has been developed in Microsoft Excel. However, although wider social costs are acknowledged, it is generally undertaken at a simple level of costs solely to the Criminal Justice System. The CSSRI, on the other hand, reflects a more comprehensive approach to costing, acknowledging, for instance, and as pertinent to this evaluation, an offender's possible take-up of health, social work and/or other community supports whilst subject to community supervision.

The Work Behaviours and Attitudes Scale (WBAS) measures individuals' attitudes and feelings towards work, co-workers, and supervisors, the extent to which they judge themselves competent in work, and the degree to which they worry about work. It was developed in the context of psychosocial rehabilitation programmes for people with long term mental health problems, motivated by the idea that interventions could be successful if they increased a person's self-esteem as a worker, or had a positive influence on work motivation or attitudes. The WBAS is based on the scale developed by Griffiths (1973), and consists of 29 questions which are scored on three sub-scales: pride and independence; work anxiety; and work performance. In addition, there is an overall score for the entire WBAS scale. The scale has mainly been used in the field of psychiatric rehabilitation, and mainly in North America (Mowbray, 1995). For its use in the INTEGRA evaluation, Cronbach's Alpha¹⁰ was calculated for the WBAS as a whole and for its three sub-scales, based on 120 cases (completions at T1 + T2): Scale overall (0.7528); Pride and Independence (0.7561); Work anxiety (0.514); Work performance (0.7132).

The Empowerment Scale was developed by researchers at the University of Boston, USA, working with a group of mental health service users (Rogers *et al*, 1997). It is a 28-item instrument which measures six dimensions of empowerment: self-efficacy, self-esteem, power, change, anger, and group/community action.

Whilst studies have demonstrated adequate validity of the scale when used with adult mental health populations in the USA (Wowra and McCarter, 1999; Corrigan *et al*, 1999), some authors argue that empowerment is contextually dependent (e.g. Zimmerman, 1995), and that the development of a universal and global measure of empowerment may be neither feasible or appropriate. It has not

¹⁰ Cronbach's Alpha is a coefficient that measures the reliability of a test, in the special sense of its internal consistency – for further consideration see Rubin and Babbie (2001) page 193.

previously been used with drug or alcohol problem users. For its use in the INTEGRA evaluation, Cronbach's Alpha was calculated for the empowerment scale overall (0.7758) and for the self-esteem dimension (0.8256) based on 120 cases.

The Maudsley Addiction Profile (MAP) was designed at the UK National Addiction Centre as a brief instrument for treatment outcome research (Marsden *et al*, 1998). It stems from the National Treatment Outcome Research Study (NTORS) (Gossop *et al*, 1998), and is now widely used and validated. It measures problems across four domains: drug use, health risk behaviour, physical and psychological health, and personal/social functioning. Based on 240 cases, Marsden *et al* (1998) report that test-retest reliability of the MAP was good: average intraclass correlation coefficients across eight substances were 0.94 and 0.81 across health risk, health problems, relationship conflict, employment and crime measures.

As a consequence of its inclusion in this evaluation, I decided to introduce a 'Consent to Disclose' (Appendix A) form for any respondent whose health caused concern at the time of interview. I would only act on those concerns with the signed consent of the respondent.

The Working Alliance Inventory (WAI) (Horvath and Greenberg, 1989) is based on Bordin's (1980) tripartite (i.e. bonds, goals, and tasks) conceptualisation of alliance, and gauges the strength of therapeutic relationship between, in our case, probationers and INTEGRA E Team Project workers. It was considered that this should be included in the core design because of the intensive case-working that the programme involved. The WAI is a 36-item measure with three sub-scales that measure client and therapist perceptions on goals, tasks, and quality of the personal bond respectively. It has two versions: the client version assesses the quality of the alliance as perceived by the client, and the therapist version assesses the therapist's perspective as to how the client perceives the quality of the alliance.

Sample items are "My counsellor perceives accurately what my goals are" (Client version) and "My client and I have a common perception of her/his goals" (Therapist version). This evaluation only employs the Client version.

Items are scored on a 7-point scale (from 1 = Rarely to 7 = Always). Horvath and Greenberg (1989) reported good internal consistency estimates for the client total score and sub-scales. For its use in the

INTEGRA evaluation, Cronbach's Alpha was calculated for the WAI overall (0.9318), based on 65 T1 completions, and for the three subscales: Bonds (0.816), Goals (0.8675), Tasks (0.7466). Validity has been established through significant correlations between WAI ratings and counselling outcome (Horvath and Greenberg, 1989), client characteristics (Kokotovic and Tracey, 1990), and therapist technical activity (Kivlighan, 1990). Kokotovic and Tracey (1990) also conducted confirmatory factor analysis and found support for a three-dimensional structure that corresponds to the three subscales.

Bearing in mind the constructivism, and 'service evaluator' (Shadish and Epstein, 1987) qualitative/formative preferences, I brought to the endeavour, this core set of INTEGRA evaluation measures was developed to include a:

Balance Sheet (Appendix B), in the respondents' own words, of the pros and cons of their contact with INTEGRA. The Balance Sheet was seen as a systematic way of gathering data related to respondents' subjective experiences of the Project which would complement the structured questionnaires (Bryman, 1988; Hammersley, 1996).

Semi-structured Interview: (Appendix C) This was specially designed for the evaluation and used at the end of the user's involvement with the Project (Time 2 [T2]) to explore respondents' views of the INTEGRA premise, and exploring themes arising from Time 1 [T1] data analysis of the cohorts. Wherever possible, this section of the interview was tape recorded to avoid the need for detailed note-taking. Themes probed include:

- respondents' understanding of, and participation in, 'informal economic' activity;
- the impact of their INTEGRA contact on their attitudes to offending and drug use;
- the impact on their work behaviours and attitudes.

These themes were conveyed using a 'report-and-response' approach (Stronach and MacLure, 1997). Essentially, this seeks to use dissemination as a tool for further data gathering, and the dissemination in this case being a summary of T1 Balance Sheets. Additionally...

- T2 interviews sought to gauge respondents' understanding of their motivation to participate in this evaluation, based on themes arising from the work of Hughes (1999).

Motivation To Participate (Appendix C)

This question of 'motivation to participate' has particular significance. As already suggested, many respondents may feel themselves to have something to risk by participating. Just why would they agree to sit down with a 'researcher' and deliver responses to the outlined array of instruments? Why, particularly, when there is no *tangible* incentive? Unlike much research undertaken by the Centre for Applied Social Studies, and despite Home Office material suggesting the contrary (in Chapman and Hough, 1998), this core design did not include *payment* for participation. For the evaluation core originator, this was attributed to 'oversight'. For the Probation Service, when asked about building in payment after my arrival, the issue was one of 'political sensitivity'. How would 'the public' feel if they heard that offenders were being paid? Clearly, the impact of even *indirect stakeholders*, as mentioned earlier in this chapter, can be highly significant - different stakes, different fears.

Fortunately, particularly given the extra demands of T2 interviews, and after having been presented with a Home Office sample cost schedule for the contracting of external evaluators (in Chapman and Hough, 1998), a schedule which included an interview payment of £10 per respondent, CDPS eventually agreed such payment for T2 interviews. Nevertheless, the significance of non-payment at T1, including ethical conundrums arising, will be discussed in more detail in the next chapter.

Semi-structured Exit Interviews – Staff (Appendix D)

As with the service users, confidential semi-structured 'exit' interviews were undertaken with INTEGRA staff. The structure and content of these interviews was agreed with E Team and Astep staff in advance to allow their comments in the light of their Project experience. The overall aim was to explore staff perceptions of the INTEGRA premise and process:

- The INTEGRA premise. Indeed, this is an area also explored with probationers, and with Case Managers in a survey of all CDPS Probation Officers – see later in this chapter.
- Perceptions of the agency partnership working.
- Perceptions of the transnational partnership working.
- Views on the value of referral criteria and process.

- Views on the scope of, and blocks to, innovation.
- Perceived Project response to personal and professional support and development.
- Balance sheet of perceived Project Positives and Negatives.

These Exit Interviews, along with the probationer T2 qualitative supplement, and the Case Manager survey mentioned above, represent emergent additions to the core design, the pseudopodia of my amoeba metaphor.

EVALUATION PROCESS

The initial design was quantitative, longitudinal, and experimental, with two cohorts of INTEGRA users making up the study group, and controls drawn from the LSI-r database of Probation Service clients in County Durham. (At my suggestion, an attempt was also made to create an experimental control of 'standard probationers' randomly drawn from the CDPS database and administer the questionnaires to them. However, this did not prove feasible as described in Chapter 6.) Each respondent would be interviewed face-to-face by the evaluator. (Initially, the design anticipated interviews taking place in probation offices; in the event it became necessary to interview respondents in their homes wherever possible.) The cohorts would be re-interviewed 12 months later and comparisons made. Cohort 1 involved sampling INTEGRA referrals for the period June– December 1998 (T1); and Cohort 2 for the period June – December 1999 (T2)). Sampling involved ALL service users in these periods who had been seen by a Project worker from the E Team for mentoring.

So, a quantitative core methodology, a 'hard-to-reach' user group, and no *tangible* incentive: so goes the conundrum at the heart of this evaluation. Just how do you go about recruiting an adequate sample ('getting the numbers'), and how do you do it ethically?

DATA COLLECTION

The Cohorts at T1

Cohort 1 involved sampling INTEGRA referrals for the period June 1998 – December 1998. Given the relatively small number of people seen by the Project during its early months, it was decided to target all those listed on the Project's database as having been seen by a Project worker for mentoring (see Chapter 3). This totalled 79 people.

The CDPS database of clients was then consulted by INTEGRA Administrative staff for information on their whereabouts, if known, and current Case Manager (Probation Officer), if still on an order. If a Case Manager was listed, they were contacted by fax, advising of my intention to contact their client with regard to the INTEGRA evaluation. Case Managers had already been alerted to my impending contact by a memo from the INTEGRA Project Manager and his line manager, the Assistant Chief Probation Officer (ACPO). This person held overall responsibility for the Project, and responsibility for chairing both the multi-agency Project Management Group and the Evaluation Management sub-Group of this. My fax to Case Managers detailed the confidentiality of the interviews, the fact that they were not compulsory, and also asked for any information which might be relevant for my health and safety.

Letters were then sent to 25 INTEGRA attenders with Case Managers, offering an appointment at a probation office. The letters outlined the evaluation and the fact that they were not compelled to be interviewed, but that confidentiality was assured if they agreed (the ethical and legal obligations of researchers regarding confidentiality in the illicit drugs field deserves more detailed consideration and will be explored in Chapter 5). Accompanying the letters were contact details for people to reply if not interested, including a Stamped Addressed Envelope (SAE) and a reply slip. My intention was to use this initial group of 25 respondents to 'pilot' the questionnaires and the interview process, then review and consider any necessary adjustments.

In the event, one person reported for interview, and there were no cancellations. I was beginning to understand what I might be up against. Although my experience of working with alcohol and other drug users, including offenders, dates back to the late 1970s, I had not been active in the field for 7 years. I needed to resensitise myself to some of the key issues, and rethink my approach. I was already being challenged about my possible 'resistance to the methodology', and urged to 'get the numbers'. But just why would drug users agree to participate in any case?

Few studies seem to have addressed the question of people's motives for participating in social research. What is reported tends to be the reflections of the researchers. As Lee (1993) points out, what is lost are the perspectives of the participants.

When discussing the concept of stakeholders earlier in this chapter, I noted that 'user participants' are potentially stakeholders too. Clearly, and perhaps not surprisingly given the enforcement imperative of Probation and the limited time available, service users were not consulted at any point in the evolution of this evaluation design. In late April 1999, I met, confidentially, with a small group of four probationers, at a probation office, to 'chat' about INTEGRA and the evaluation. Two were active and regular drug users – it emerged.

This was a 'meeting of two halves'. The first was about establishing a rapport. The second half was spent using this rapport to get the attenders' views on some key issues. It was during the second part of the meeting that those present felt able to begin disclosing, particularly about their offending, their drug use, and their attitudes towards employment and training opportunities.

Key issues which emerged were:

- The two drug users felt uncomfortable about discussing their drug use in a probation office. They needed repeated reassurance about confidentiality.
- The two drug users recognised a connection between their offending and their drug use.
- The three unemployed members of the group felt 'justified' in their use of the 'informal economy' to boost benefit income.
- If I was standing on their doorstep, and they did not know who I was, they would have 'second thoughts' about opening the door to me, "you look like a debt collector or something". (At that point in the evaluation I was in the habit of 'cold calling' on addresses where I had previously found no-one at home for an arranged interview. The group also thought that cold-calling was "not very respectful".)
- All group members would feel uncomfortable about participating in the INTEGRA research, especially if they had to see me as part of their Probation Order. Having met me, though, and having had it explained to them, especially once 'feeling' that it was confidential, they would be more inclined to say "yes", so long as it had some relevance for them, for example, if they had attended the programme and could 'see the sense'.

Of course, different ethical and practical issues emerge when recruiting particular groups of people for research purposes (Homan, 1991). For instance, Winckler's (1987) reflections on

company directors being made to feel "special" when invited to talk about their experiences might hold little relevance for drug using offenders wary of prying eyes and the risks of exposure (Brain *et al*, 1998). So what might hold relevance for them?

In a qualitative study of drug injectors (n=17), Hughes (1999) reports a variety of reasons for participation. Themes include:

- Interesting, relevant, and worthwhile research;
- The timing of the fieldwork and respondents' priorities;
- The benefits of the research to respondents;
- Research benefits to people in general;
- The impact of referrers on participation;
- The respondents' first impressions of the researcher;
- The importance of confidentiality;
- The use of tangible rewards and incentives.

Of particular significance for my attempts to recruit participants for this evaluation seemed to be:

- One form of non-participation regularly encountered by Hughes was people failing to keep appointments for pre-arranged interviews. Such non-appearance might constitute an indirect refusal to participate. But, as noted earlier in this chapter, it might also signify that set interview times do not fit comfortably with aspects of a drug user's lifestyle. Rather than interviews in an office, then, *home visits* might be more successful.
- Hughes found evidence that access to respondents relied on some degree of trust being established. For this to happen in my case, not having met the people previously, the intervention of a third party would be necessary. This 'gatekeeper' could be anyone who already commanded some trust, and could vouch for me. A project worker perhaps, or even someone with more personal contact – mothers, lovers, or others – people I could meet and engage on my home visiting trails.
- The first impressions made by a researcher can have a marked impact on people's willingness to participate. This, of course, could be face-to-face and include verbal cues, or simply while standing on a doorstep having my appearance judged from behind a twitching curtain. Power (1989), for example, remarks that informal, casual dress is preferable when meeting drug users. Expensive clothes can cause

unease and embarrassment, and can also suggest 'authority'. And as for verbal cues, Hughes found that: 'a non-judgemental attitude on the part of the interviewer together with the prospect of a 'no ties' interview are key features that encourage participation' (Hughes, 1999 p. 321).

- Hughes also notes that a small *material incentive* is a practical way to encourage people to participate: 'As with all reasons for participating it is important to recognise that the use of incentives in social research brings with it a number of ethical and practical considerations' (p. 322). Such ethical considerations are explored in Chapter 5.

It is interesting to note another of Hughes' caveats at this point.

Participants offered their accounts in a face to face situation with a researcher who had sometimes both recruited and interviewed people. Some individual participants spent a considerable amount of time with the researcher, often meeting regularly over a period of weeks and months. Having invested this time, participants may feel a need to justify participation to themselves or the researcher. These factors could help to account for the positive and broadly altruistic responses gleaned in this study...it is interesting to note that some people did comment on their participation in socially 'undesirable' ways, such as when people reported the influence of financial incentives and rewards. (ibid, p. 322)

I should say, too, that after years of working in the drugs field, none of this was 'news' for me. However, finding Hughes' work was a useful exercise in reacquainting myself with some issues.

I should also say, though, in the absence of *tangible* incentive at T1, I found it necessary to apply some 'gentle coercion'; that statement does need some immediate ethical clarification.

Rather than spelling out the notion of 'informed consent' in my appointment letters, I decided to leave this until I met face-to-face with the person. In arranging home visits too, I decided on an 'opt-out' rather than 'opt-in' approach – I would call at their home at 'such and such' a time, on 'such and such' a day, unless I heard from them otherwise. The letters still included full contact details, including a SAE and a reply slip, but were different in tone to the original. In any case, any subsequent interview would still only

proceed with their signed consent (Appendix A). Nevertheless, I was definitely more 'pushy'. As explained in the next chapter, I felt ethically justified in this. There seemed little chance of doing 'harm' to the recipients of my letter, and whatever the ethical price of proceeding this way, this had to be balanced against the ethical price of not getting to see sufficient numbers to validate the evaluation, thereby undermining it, and undermining the 'good' that could come from its use.

This combination of approaches, with a particular emphasis on home visits, together with a dose of tenacity and rigour in keeping on top of repeat visits and letters, telephone calls, changes of address, and contacts with 'others', did produce some improvement in recruitment. However, it was very time consuming.

It seems fair to say that people were generally positive about my visits, once I was able to meet and talk, with only a handful refusing consent face-to-face. Some did confess to finding my contact 'pushiness' a bit 'cheeky', but were still willing to participate. Indeed, many of those interviewed seemed genuinely interested in the evaluation, and the individual questionnaires which were explained to them. Many also seemed to enjoy the chance to consider how they 'think and feel', a relatively novel experience for them it seemed. So, perhaps Winckler's (1987) company directors are not so different to INTEGRA's unemployed drug using offenders after all.

I believe this generally positive reception not only raises important questions around interview approach, but also around...

...interview delivery

As Krosnick (1999) notes:

One prevailing principle of the survey method is that the same questionnaire should be administered identically to all respondents (e.g. Fowler and Mangione, 1990). If questions are worded or delivered differently to different people, then researchers cannot be certain about whether differences between the answers are due to real differences between the respondents or are due to the differential measurement techniques employed.

If respondents expressed uncertainty and asked for help, interviewers avoided interference by saying something like 'it means whatever it means to you'. Some critics have charged that this approach compromises data quality instead of enhancing it. (p. 5)

Such critics (for example: Briggs, 1986; Mishler, 1986; Suchman & Jordan, 1990, 1992.) argue that the meanings of many questions are inherently ambiguous, and that their meaning should be 'negotiated' in 'everyday' conversation and back-and-forth exchanges between the researcher and respondent. To prohibit such exchanges is to straight-jacket them and prevent what is needed to maximise response validity. Recent work by Schober and Conrad (1997) would seem to verify such argument. They claimed that when researchers were free to clarify the meanings of questions and response choices, the validity of reports increased substantially.

Although I initially felt pressure from the traditional viewpoint to resist clarification and exchange about the instruments, this was modified by the 'pilot' experience of the first few interviews, and such exchange became comfortable and commonplace. I believe such exchanges, too, helped in establishing the necessary rapport for meaningful exchanges during the qualitative elements of the interviews which occurred towards the end.

Despite a generally positive response to the interviews, though, recruitment remained an enduring problem throughout this evaluation. This was raised at a meeting of the Probation-chaired Evaluation Management Group on 12th April, 1999. At that meeting, it was suggested by CDPS that 'research' contacts could be made 'reporting' contacts under National Standards (Home Office, 2000). This would mean that failure to attend interview, or not be at home when I called, would be recorded by the case manager as an unexplained absence. Two such absences usually result in a return to court for Breach proceedings, a further sentence, and possible imprisonment. After careful ethical consideration, including the meeting with a group of probationers mentioned earlier, and despite the differing views of my manager, I decided not to go down this route. The ethical issues involved in this decision, along with possible implications arising from the Human Rights Act 1998, are discussed in detail in the next chapter.

EMERGENCE AND THE PSEUDOPODIC FRINGE

In line with my utilisation-focused (Patton, 1997) commitment to be active-reactive-adaptive, and in addition to the already mentioned semi-structured qualitative additions to the methodology, including my exploration of respondent 'motivation to participate', other areas emerged during the course of this evaluation which, I believed, demanded some systematic investigation. These further illustrate the pseudopodic fringe of my amoeba metaphor:

The Case Managers (Appendix E)

As with similar studies (e.g. Sarno *et al*, 2000; Edmunds *et al*, 1998), I took the view that Case Managers (Probation Officers) are major stakeholders in any initiative undertaken by CDPS, and any evaluation of the kind undertaken would be incomplete without their feedback. This had also been a view expressed by E Team Project staff who felt that they would appreciate systematic feedback from Case Managers. This addition to the evaluation was approved by the Research Management Group at a meeting on 1st July 1999. However, it was felt that distribution of the questionnaire should wait for a few months. Probation Officers, I was informed, had just undergone a Home Office inspection and were somewhat 'sick' of questionnaires at that point. As such, in October 1999, 15 months into the INTEGRA lifespan, a confidential and anonymous questionnaire, along with SAE and my contact details for any queries, was distributed to all relevant CDPS Probation Officers via their location managers. This distribution was preceded by a memo from the Assistant Chief Probation Officer to all recipients and their managers encouraging completion of the questionnaire.

This questionnaire, which had previously been approved by INTEGRA Project staff and the multi-agency Project Management Group, sought to explore Case Manager agreement, or otherwise, on a 5 point Likert Scale, with 18 statements about offender employment interventions in general, and about the INTEGRA Project in particular. These statements covered issues already raised anecdotally by some Probation staff in conversation with myself, and some issues addressed by the Sarno *et al* (1998) study:

- Project premise.
- Role of the Probation Service in employment intervention.
- INTEGRA impact on existing agency support and referral network.
- Clarity regarding INTEGRA objectives and working methods.
- Satisfaction with the work of the Project and its staff.

The precise statements are listed in Chapter 6, along with a chart of all responses. Taking into account staff movement within the service, and staff departures and arrivals, the 30, mainly anonymous, returns represent around a 100% response as far as I can ascertain.

Drug Use of People in Cohorts But Not Seen

Developing the notion of respondent contactability bias, I decided to explore the drug use of those people included in our cohorts and who had been assessed by the Project, but whom I was unable to 'capture' for evaluation interviews. I hypothesised that their use of drugs other than alcohol and/or cannabis would be greater than those interviewed. This was achieved with a systematic trawl through case files. At the same time, I extended my investigation of the CDPS LSI-r database (the source of control data) to also explore the same cohort individuals who had been assessed but not interviewed for the evaluation. In this way, data analysis would be able to further explore differences between those interviewed and those not.

DATA ANALYSIS

The Centre for Applied Social Studies uses Formic™. This is a software tool for the design and processing of surveys. It provides facilities for:

- The design and production of forms
- The scanned capture of data from completed survey forms
- Processing and exporting the retrieved data

At the outset of this evaluation I designed/converted all instruments into Formic™ format, which were then printed for use with respondents. Completed forms were scanned into Formic™, and then exported into SPSS™ for data checking, editing, and analysis. Likewise, data from the Project database (in Microsoft Access format) and from the CDPS LSI-r database (in Microsoft Excel format) were also converted/imported into SPSS™. The Case Manager survey was designed from scratch in Formic™. My involvement in the analysis of quantitative data was restricted to checking, editing, and the production of descriptive statistics where applicable. More detailed statistical analysis was undertaken by the evaluation manager, Dr. Justine Schneider, and by Dr. Pauline

Coolen-Schrijner from the Statistics Consultancy Unit at the University of Durham.

To manage the processing demands of the semi-structured qualitative elements of this evaluation, including the Balance Sheets, I used CDC EZText™. This is a public domain software package specifically designed by Conwal Inc. and the Centers for Disease Control and Prevention (CDC) in Atlanta to assist researchers create, manage, and analyse semi-structured qualitative databases (Carey *et al*, 1998). In this way, I designed data entry templates for each component series, and typed respondent-identified data directly into these – either from hand written hard copies completed during the interviews, or from tape-recordings where respondents permitted their use. Following data entry, I created online codebooks, and applied codes to specific response passages. Relevant text passages and numbers of respondents/responses were then identified by conducting database searches according to these codes. For the Balance Sheets, I concluded by exporting numbers of responses based on codes into Microsoft Excel to create frequency tables.

DISSEMINATION AND USE

As Patton (1997) points out:

In utilization-focused evaluation, use does not center on the final report. Traditionally, evaluators and users have viewed the final report as the climax...and the key mechanism for use. From an academic perspective, use is achieved through dissemination of a published report. Moreover, use often doesn't emerge as an issue until there is something concrete (a report) to use. By contrast, utilization-focused evaluation is concerned with use from the beginning, and a final written report is only one of many mechanisms for facilitating use. (p. 329)

Recognising 'use' and dissemination as a process, then, I was conscious throughout this evaluation of the need to provide suitable feedback, and encouragement for use, on a continual basis. As such my intention, in addition to regularly attending Project and Evaluation Management Group meetings, was to keep in close contact with all aspects of the Project and its personnel, to create a website, and to submit findings and issues to whatever local and

national forums presented themselves. A summary of dissemination and 'use' activities undertaken is listed in Chapter 6.

SUMMARY

At its core, the INTEGRA evaluation design was quantitative, longitudinal, experimental, and summative. This is not surprising bearing in mind the Project's criminal justice heritage, and the necessary evaluation questions derived from the 'What Works' imperative of its primary intended user – the Probation Service. However, the Project's European Social Fund heritage also brought a demand for it to be innovative. The evaluation design, then, also sought to develop a flexible and formative character, allowing it to address other questions which could emerge as both the evaluation and the Project evolved and tried to innovate. One such question concerned respondent motivation to participate.

A key feature of this evaluation was service users failing to keep appointments for pre-arranged interviews. This was a particular concern given the quantitative nature of the core design, and the need to recruit sufficient numbers to satisfy statistical analysis requirements. To help overcome this problem, the methodology switched from the original intention of interviewing respondents in probation offices, to undertaking home visits. Nevertheless, understanding respondent motivation to participate was regarded as an important addition to the evaluation, not least in exploring the notion of 'contactability bias'.

Other additions to the core design resulted in a survey of Case Managers, and in developing qualitative components for data gathering and analysis.

My methodology employed to gauge primary intended user response to the evaluation and its report will be covered separately in Chapter 7.

CONCLUSION

Crafting a mixed-method evaluation from an experimental and summative core was not without some discomfort. This, I feel, occurred mainly at the level of paradigms.

The core design originator, and the contract holder for this evaluation, had a background in academic evaluation research, and a philosophical leaning towards positivism. This, I believe, was

particularly helpful for this evaluation in meeting its 'What Works' demands. On the other hand, I brought a *service evaluator's* preference for assisting programme improvement and decision-making, and formative approaches (Shadish and Epstein, 1987). I also brought, what at the outset was unarticulated, constructivist hesitation.

Whilst blending the broad positivism of my manager with the broad constructivism of myself was made easier by a shared commitment to utilisation-focused evaluation approaches, our differing philosophies, at times, resulted in disagreements over methodology and, as will be explored in the next chapter, ethics. However, in hindsight, I also believe that this evaluation benefited from the additional effort and thought that was required to overcome such disagreements.

Chapter 5

NAVIGATING THE ETHICAL SPACE

Without ethics, everything happens as if we were all five billion passengers on a big machinery and nobody is driving the machinery. And it's going faster and faster, but we don't know where.

Jacques Cousteau, Ted Turner television interview, CNN, 24 February 1989

INTRODUCTION

Following Kushner (2000), evaluation creates an 'ethical space' (a term coined by Poole, 1972):

This is a space in which some people are invited to make novel judgements about the work of others, and in which the nature of those novel judgements can be regulated and scrutinized...through agreement on principles and procedures governing the conduct of the evaluation, ruling on participants rights and the obligations of the evaluation in respect of them. (Kushner, 2000, p.151)

In navigating this ethical space, most evaluators, as with social researchers generally, rely on four sources as guides for their ethical thinking (Newman and Brown, 1996, p.3). Three of these: *intuition, past experience, and observations of colleagues' conventions*, rely on the evaluator's learning, past and present, and, clearly, are very appropriate so long as they are not applied haphazardly. The fourth source, *statements of ethical or professional rules and codes*, again are very appropriate, and, indeed, are the hallmark of maturity for any profession. Once again, though, they are insufficient if used indiscriminately, or if the evaluator has no experience of relating these rules or codes to broader ethical theories and principles to assist their thinking.

This chapter seeks to clarify how I navigated the ethical space within the INTEGRA evaluation, drawing on the 'Guiding Principles for Evaluators' of the American Evaluation Association⁹, and employing Newman and Brown's (1996) 'cognitive process' framework for ethical decision-making. In so doing, I will elaborate on a number of specific ethical dilemmas, some of which were touched upon in the previous chapter: 'voluntary informed consent', 'privacy', and 'consent to disclose'; 'disclosure after death'; 'coercion as incentive'; and 'payments for participation'.

This chapter could be synthesised with the previous. However, I consider the themes outlined here to be so significant as to warrant their own distinct consideration. This reflects my belief that ethical thinking is too often limited to, and too often suppressed by, code-driven design concerns.

Before proceeding to INTEGRA ethical decision-making specifics, I will begin by providing some definitions and theoretical background, and by introducing Newman and Brown's (1996) decision-making framework.

RULES AND CODES, THEORIES AND PRINCIPLES

DEFINITIONS

Table 5.1: Distinctions Between Ethical Rules, Codes, Principles, and Theories

RULES	Statements of specific dos and don'ts.
CODES	Compilations of rules, usually adopted and endorsed by a professional organisation.
PRINCIPLES	Broader than rules or codes; provide guidance when rules conflict or when rules are not specific to the context.
THEORIES	Justification or criteria for ethical decisions; the science and rationale for making ethical decisions.

(Adapted from Newman and Brown, 1996, p.23)

⁹ These principles were selected as a broad guide in the absence, then, of a set of guiding principles written in a UK context. The UK Evaluation Society has now generated a set of guidelines to help commissioners, practitioners, and participants establish good practice in the conduct and dissemination of evaluations (2002) (Appendix G).

Acting ethically is not simple, and it does not render professional life any easier, indeed it makes it more complicated. However, no one welcomes their work being considered unethical. Understandably, some defensiveness can form a part of our response to any queries regarding our commitment to ethical practice. A common repost to such queries, according to Newman and Brown's research, is:

The answers to ethical dilemmas lie in our professional ethical codes. (p.186)

However, the thrust of Newman and Brown's argument is that ethical rules and codes, on their own, and whilst potentially helpful, are often insufficient for the task of ethical decision-making, and any sole reliance on them merely represents an ethical minimalism. Indeed, Pattison (1999), writing from a UK counselling perspective, argues that unswerving reliance on professional and ethical codes may be unethical in itself as it cages and disempowers informed ethical judgement. Rather, then, we need to embrace broader ethical principles and theories if we are conscientiously to commit to responsive and thoughtful ethical practice.

ETHICAL THEORIES

To group, summarise, and compare the plethora of ethical theories, Newman and Brown adopt a five category classification to help consider what useful ethical guidelines each may hold for evaluators. These categories are: *consequences*, *duty*, *rights*, *social justice*, and *ethics of care*. Whilst these are not necessarily mutually exclusive, nor do they necessarily lead to different ethical decisions, they are, nevertheless, useful organisers to consider the rationale for making ethical choices. Summarising each in turn:

1). Consequences

These are theories which propose that the rightness or wrongness of an action is determined by the consequences of the action. The most prominent is *utilitarianism*. This views actions as right when they result in the greatest possible good for the greatest number of people. To take the right action, you need to estimate the effect of each action on all the parties involved and select the one which optimises the satisfaction of the greatest number. Utilitarian theories are further categorised as either *act* or *rule* utilitarian:

- **Act utilitarian** stances emphasise making decisions solely on the outcomes or consequences of an act. Act utilitarians are more likely to make a decision based on the particulars of a

specific situation rather than search for a rule applying to all similar situations.

- **Rule utilitarians**, on the other hand, would search for a relevant rule and follow it. Rules, according to this perspective, must be followed consistently and not be twisted to fit particular circumstances. This might not lead to the greatest good in all situations, but, over the long term, rule utilitarians believe that this approach will result in the greatest societal benefit compared to all alternatives.

When using utilitarian ethics and the criteria of consequences as a guide for ethical decision-making, two questions are particularly salient for the programme evaluator:

- In seeking to maximise the good for the greatest number, how is 'good' defined and determined? What does 'good' mean?
- Whose 'good' is it that matters? What if those receiving a service, for instance, benefit but at the expense of those providing the service? With the INTEGRA Project in mind, for example, what if the service users succeed in overcoming their drug difficulties and securing sound employment, but the project staff burn out and leave the field, never to offer others the benefit of their skills and experience?

2). Duty

Deontologists assert that right actions are not determined by their consequences but, rather, by our sense of duty based on our relationships and other criteria, such as personal commitment, the existence of a rule or law regarding behaviour, or perhaps a religious command. Kant is most associated with this theory in the modern era through, what is known as, his *categorical imperative*. According to Kant (1788/1956), all people should act as if their actions were to become universal. When making a decision about how to act, we must ask what would happen if everyone acted in similar circumstances in the same manner. For instance, an evaluator might be tempted to 'soften' the reporting of a negative finding because the programme manager is conscientious, caring, and already under an unusual amount of stress. No apparent harm would result. But what if every evaluator did the same? What would be the result for the programmes they evaluate, or the impact on broader trust and faith in evaluation efforts?

3). Rights

Unlike a strict utilitarian who might justify an infringement of individual rights for the betterment of total society, a supporter of individual freedom would find such behaviour inappropriate. Most ethicists discuss rights as claims that individuals and groups can make on others. A right, then, can be viewed as property one possesses and over which one has control. Most philosophers, though, would not consider rights as absolute. Conflicts may arise between rights, and decisions must be made about which rights, or whose rights, take precedence. Evaluators, for instance, respect a person's right to privacy regarding information revealed. But if harm to someone, say, is a possible outcome, then serious consideration must be given to infringing the right to privacy.

The notion of 'infringing' a right refers to a justified action overriding a right (Beauchamp and Childress, 2001). A 'violation' of a right, on the other hand, refers to an unjustified action against a right. Decisions regarding justification to infringe are not straightforward. At the local level, evaluators might find it useful to examine their decision from the perspective of consequences, and weigh the relative balance of infringing or not. At the broader societal level, the tension between individual rights and social responsibilities lies at the heart of discussion and legislation regarding citizenship and participation. Etzioni (1999) argues that the right to privacy in contemporary society may have reached a point where privacy is becoming counter-productive, and that we may need to create a more 'transparent' society. His notion of *communitarianism* (Etzioni, 1996) holds that a good society must seek a carefully crafted balance between individual rights and social responsibilities, between liberty and the common good. As a way of crafting this balance, he proposes a formula of criteria to be met when privacy may have to give way to the greater public good:

- When there is a clear and present danger to society;
- When there is no other way;
- When the intrusion is minimal;
- When measures are in place to remove undesirable side effects.

Engaged in a 'War Against Terrorism', the first of these criteria currently underpins a clutch of rights-restrictive counter-terrorist legislation on both sides of the Atlantic. But even beyond the special circumstances following the tragic events in New York and Washington on September 11th, 2001, there may be a case for greater limits to privacy. However, as others caution:

...it is also vital that the measures in place to counter undesirable effects are in reality strong enough to withstand the danger of incremental moves towards totalitarianism. (Rice and Thomas, 2001, p.29)

In the UK, the Human Rights Act 1998 (HRA), which came into force on 2nd October 2000, can be seen as one such rights-protection measure, and represents a New Labour commitment to rights-based law by bringing much of the European Convention on Human Rights into domestic courts. The intention of the HRA is to encourage all public authorities and their contracted agents, which can be taken to include external evaluators, to reconsider their practices from a human rights perspective. Further mention will be made of the HRA later in this chapter.

4). Social Justice

Rawls (1971) is one of the most prominent theorists advocating an egalitarian, social justice perspective. He emphasises equity, but adds a dimension that is in opposition to a utilitarian stance by asserting that efforts to combat social and economic inequalities should be arranged to benefit the least advantaged. House, as outlined in Chapter 2, has been the most prominent voice in evaluation suggesting that evaluators embrace social justice issues and consider Rawls' (1971) approach. For House, evaluators need to become advocates for the disadvantaged and powerless. Even when a stakeholder approach to evaluation is adopted, the powerless are seldom involved and, too often, the critical issues are not discussed:

Evaluators cannot be value neutral in these matters. Our conceptions and even our methodologies are value laden. Evaluators do not live in a state of methodological grace. (House, 1991, p.245)

House's notion of justice finds its most practical application in his arguments for a fair and demanding evaluation agreement (House, 1980), in which all participants should meet the demanding conditions that they:

- Not be coerced;
- Be able to argue their position;
- Accept the terms under which the agreement is reached;
- Negotiate;
- Not pay excessive attention to one's own interests;
- Adopt an agreement that affects all equally;

- Select a policy for evaluation that is in the interests of the group to which it applies;
- Have equal and full information on relevant facts;
- Avoid undue risk to participants arising from incompetent and arbitrary evaluations.

5). Ethics of Care

The four categories of ethical theories covered so far, focus on logic and reasoning. Some feminist perspectives on ethics, however, give greater attention to relationships and the specific context of the ethical dilemma. Gilligan (1982) believes that women are more likely than men to consider how an ethical decision will affect human relationships and more likely to seek a resolution that fits the specifics of the situation. An ethics of care, which she proposes, would be more concerned with relationships and context than with universal laws or principles. Noddings (1984) suggests that an ethics of care is not a rejection of rational thinking, but that rather the rational thinking must serve something higher. And that something is caring for others. Typically, then, proponents of an ethics of care emphasise the importance of considering how particular individuals and organisations would be affected by available choices.

An ethics of care is congruent with constructivist perspectives generally, and with those that emphasise empowering the disenfranchised particularly. Clearly there is much overlap with social justice standpoints, and evaluators who favour formative approaches, assisting programmes in a nurturing and supportive way, would also find an ethics of care consistent with that role.

Usefulness of theories

Each of the theory categories outlined provide a framework for ethical decision-making. Refocusing each as questions:

- 1). What are the consequences of my decision?
- 2). What duties and obligations do I have as an evaluator?
- 3). What rights do respondents have? What rights do I have?
- 4). What would be just or fair in this situation?
- 5). What would be the caring response?

Each question can be asked as a way to help us consider ethical choices. As will be outlined later when considering Newman and Brown's decision-making framework, the relative importance of each question may well reflect our personal values and beliefs as well as our philosophical preferences as evaluators. For instance,

an ethics of care stance has the advantage of considering specific contextual factors, but may be too relativist for evaluators seeking universal rules, or for those who advocate a more positivist and distanced stance.

Ethical theories lead to the development of ethical principles, which in turn can serve as a foundation for establishing ethical rules and as the organisational framework for ethical codes. The process of developing ethical codes (i.e. from theories, to principles, to rules/codes) is not necessarily so clear and linear. But, according to Newman and Brown:

...it is helpful to understand the possible relationships when trying to make ethical decisions. (p.37)

ETHICAL PRINCIPLES

Various principles and models could be used as a framework for evaluation. House (1993), for instance, suggests three principles:

...mutual respect, noncoercion and nonmanipulation, and support for democratic values and institutions. (p.167)

Kitchener (1984), drawing heavily on the biomedical work of Beauchamp and Childress (1983 – see 2001 for latest edition), presents five principles as specifically relevant for the helping professions: *respecting autonomy*, *nonmalificence* (avoiding harm), *beneficence* (doing good), *justice* (being fair), and *fidelity* (being faithful). Once again, as with theories, each of these can be refocused as example questions that may help us consider ethical choices:

1). Respecting Autonomy

egs. Are anyone's rights affected, including the evaluator's? Are the rights of respondents being respected?

2). Avoiding Harm

egs. What harm is likely to arise as a consequence of a decision? Will staff be exposed to excessive stress because of the evaluation process? Will respondents be harmed in some way?

3). Doing Good

egs. What good will come, and for whom, through the evaluation? Is the maximum good being achieved? What good can be

accomplished beyond the expectations of professional codes and rules?

4). Being Fair

egs. What issues are related to fairness and accuracy in this evaluation? Are multiple perspectives being gathered?

5). Being Faithful

egs. What contractual arrangements have been made? Do service users expect the evaluation to advocate on their behalf?

I have to say, at this point, and in my opinion, that the principle of 'avoiding harm' takes precedence over all others. This, though, like any principle, cannot be an absolute. It does not mean, for instance, ignoring negative findings, however harmful these may be for some. The emphasis is on 'undue' harm. Negative findings, then, should be presented in a way that strives to avoid 'undue' harm to the programme, the staff, or the service users. In other circumstances, risking, or even inflicting, some harm may be inescapable to avoid an even greater harm, or necessary if we are to advance some greater good. At such times, though, evaluators need to ensure that they have fully considered their decision, and that the principle of 'respecting autonomy' means that those on the 'receiving end' are at least fully informed of, and ideally fully involved in, the decision-making.

The Guiding Principles for Evaluators, adopted by the American Evaluation Association (AEA) and which I used as a foundation for my approach to the INTEGRA evaluation (see Chapter 4), implicitly include Kitchener's (1984) principles. Although the AEA statement makes no explicit mention of such broader ethical principles, Table 5.2 below summarises how they can be related to Kitchener's formulation:

Table 5.2: Linking the AEA Guiding Principles to Kitchener's Ethical Principles

AEA Guiding Principle	Kitchener's Principle
<p>Systematic Enquiry Evaluators conduct systematic, data-based inquiries about what is being evaluated.</p>	<p>Doing good. Avoiding harm. Being fair.</p>
<p>Competence Evaluators provide competent performance to stakeholders.</p>	<p>Avoiding harm.</p>
<p>Integrity/Honesty Evaluators ensure the honesty and integrity of the entire evaluation process.</p>	<p>Being faithful.</p>
<p>Respect for People Evaluators respect the security, dignity, and self-worth of the respondents, programme participants, clients, and other stakeholders with whom they interact.</p>	<p>Respecting autonomy. Doing good.</p>
<p>Responsibilities for General & Public Welfare Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare</p>	<p>Being fair.</p>

(Adapted from Newman and Brown, 1996, p.53)

Similarly, Kitchener's principles can also be related to the ethical theories outlined earlier:

Table 5.3: Linking Kitchener's Ethical Principles to Ethical Theories

Theory Category	Kitchener's Principle
Consequences	Respecting autonomy. Avoiding harm. Doing good.
Duty	Respecting autonomy. Avoiding harm. Doing good. Being faithful.
Rights	Respecting autonomy. Avoiding harm. Being fair.
Social Justice	Being fair. Avoiding harm. Doing good.
Ethics of Care	Avoiding harm. Doing good.

(Adapted from Newman and Brown, 1996, p.54)

The principles and the theory are highly interactive, and, like rules, they cannot always be applied in a straightforward way. Sometimes they will conflict. When ethical rules conflict, we consider ethical principles for guidance. When the principles conflict, or do not provide a clear picture, we use the criteria derived from ethical theory as more abstract guides. This then is the logic at the heart of Newman and Brown's framework for ethical decision-making:

NEWMAN AND BROWN'S (1996) FRAMEWORK

According to Newman and Brown, there are five levels of ethical thinking:

- Level 1. Intuition.
- Level 2. Rules and codes.
- Level 3. Principles and theories.
- Level 4. Beliefs, values, and philosophy.
- Level 5. Planning for action.

Summarising each in turn:

Level 1. Intuition

This level occurs as our first reaction to a situation prior to being reflective. At this stage we sense that something is 'not quite right'. We react on the basis of our prior ethical knowledge and experience, on the basis of our 'educated intuition' (Newman and Brown, 1996, p.93). Being alert to our intuition is particularly important when there is no time for reflection or consultation. Kitchener (1984) suggests that to act against our ethical intuition is more likely to lead to an error.

Level 2. Rules and Codes

As helpful as an educated or practiced intuition might be, it is also insufficient. This leads us to the next level of ethical thinking: reflecting and considering resources that might help. Ethical rules and codes are available for most professions. In the UK, however, until recently, and at the time of undertaking this evaluation, there were no rules or codes specifically for social work/care research/evaluation. Instead, we drew on codes available in other disciplines such as sociology or psychology. These cover issues such as: confidentiality, informed consent, seeking information sensitively, avoiding harm, recognising bias, and responsible reporting.

In April 2002, the British Association of Social Workers launched its new Code of Ethics (BASW, 2002) at the Association's Annual Study Conference in Brighton. This code now includes a section on responsibilities in the research setting (s. 4.4.4) (Appendix F), and stems from the work of the 'Theorising Social Work Research' seminar series held between May 1999 and December 2000. Funded by the Economic and Social Research Council, and organised by the Joint University Council Social Work Education Committee, the National Institute for Social Work, and the Association of University Professors of Social Work, the seminar series produced a draft code of ethics. Following wider consultation, and some amendment, this code has now been adopted.

Following Newman and Brown's framework, then, rules and codes can offer useful guidance if we can find one that fits our situation. If we do find a fit, we have three choices: take immediate action (Level 5), move to the principles and theory level (Level 3) for further analysis, or do nothing. If we do not find a fit, then we move to Level 3:

Level 3. Principles and Theory

At this level, Newman and Brown suggest we first analyse the situation from each of Kitchener's (1984) ethical principles: 'respecting autonomy', 'avoiding harm', 'doing good', 'being fair', 'being faithful'. We should ask what is the relevance of each principle, and if principles conflict, how might they be balanced? At this point, even if we are close to taking some form of action, Newman and Brown advise we consider the relevance of criteria and questions derived from ethical theory: 'consequences', 'duty', 'rights', 'social justice', and 'ethics of care'. This is particularly helpful if the principles do not provide sufficient guidance or if they conflict.

Level 4. Beliefs, Values, and Philosophy

The fourth level of ethical decision-making, suggested by Drane's (1982) work, emphasises the importance and consideration of our beliefs, values, and philosophy. Drane suggests that the first level of ethical thinking, the intuitive, is closely related to this level four where we consciously attempt to use knowledge of our values and beliefs. Both levels represent what it is we truly believe, what it is we value, and what kind of person or professional we want to be. At the fourth level, though, we make these explicit. For instance, each of us will no doubt hold a felt hierarchy of ethical principles. As mentioned earlier, my personal paramount is 'avoiding harm'. This, then, will have a significant bearing on whatever ethical decision I face. It owes much, I believe, to aspects of my philosophical centre of gravity – as outlined in Chapter 2.

Given my constructivist hesitation regarding reality and knowledge, I believe myself to be less inclined towards any 'justified' infringement of respondents' rights, and more inclined to classify any potential harm as 'undue' than, say, some (post)positivistic empiricists 'rigorously' searching for, what I believe to be, the naïve realism of 'empirical truth', and holding, what I again believe to be, naïve expectations of 'the good' that could come from the search. Similarly, I believe myself to be more inclined towards ethical decisions which embrace 'social justice' and 'ethics of care' theoretical perspectives. There are some things, quite frankly, which are fundamentally more important than some of our asinine ends and some of our asinine means for achieving them. Whilst in practice this may mean that, on occasion, I may have different ethical thresholds than some colleagues, it does not mean that I necessarily consider myself to be a, somehow, more ethical practitioner – but merely that I see the ethics differently, I see the reality(ies) differently, and I see the role(s) of evaluator differently.

Level 5. Action

At this fifth level, which Newman and Brown have added to those of Drane (1982) and Kitchener (1984), we move from thinking and feeling to acting on our ethical belief. Newman and Brown's aim is to fill, what they feel to be, an ethical action process vacuum:

What courses of action are open to us? Whom should we consult? What should happen next? Discussions of what to do once we find a rule or standard violated or in danger of being violated are sparse. (Newman and Brown, 1996, p.95)

Newman and Brown suggest six action steps:

1. Analyse how much stress is involved for ourselves and others.
2. Consult with colleagues.
3. Consider what impact the programme climate, our own organisational situation, and any cultural issues should have on our decision.
4. Design a course of action.
5. Implement the plan.
6. Assess the impact of the plan.

All of which appears very neat and straightforward. But it is important to remember that ethical decision-making often occurs in a muddled and emotional context where there is conflict and where some confrontation may be necessary. Such conflict might be internal as we struggle to find the right course of action, but it may also be external, between our views and those of stakeholders, say, or between our views and those of our supervisors and managers. Undoubtedly, making ethical decisions can be stressful and can sometimes involve some degree of professional risk. Newman and Brown's framework does not remove the stress or the risk, but by infusing the cognitive into the emotional it can help us navigate the ethical space with more active confidence and candour.

THE INTEGRA ETHICAL SPACE

Whilst the following is inevitably retrospective, it aims to illuminate how I actively sought to 'apply' ethical thinking to specific aspects of my participation in the INTEGRA evaluation:

VOLUNTARY INFORMED CONSENT AND PRIVACY

It is a general ethical principle that you do not seek to evaluate, or have take part in an evaluation, anyone who does not know they are involved or who does not wish to take part. This is usually referred to as the principle of *voluntary informed consent*. (Robson, 2000, p.29)

Whilst there are some arguments to justify deception in social research generally (Clarke, 1999), these have 'little or no part to play in evaluation and certainly should not be used to trick people into being involved when if they did not know what was really involved they would have been unlikely to take part' (Robson, 2000, p.32).

As a way of formalising the principle of voluntary informed consent, I designed two consent forms: one for the first interview which included consent to contact respondents again in 12 months, and one for the second interview which included an acknowledgement that this interview concluded respondents' participation in the evaluation. Both forms also emphasised that respondents could withdraw their consent at any time and without prejudice to themselves in any way.

However, as mentioned in Chapter 4, rather than spelling out the notion of 'informed consent' in my appointment letters, I decided to leave this until I met face-to-face with the person. In arranging home visits too, I decided on an 'opt-out' rather than 'opt-in' approach – I would call at their home at 'such and such' a time, on 'such and such' a day, unless I heard from them otherwise. The letters outlined the reason for my visit, and included full contact details, including SAE and a reply slip. The reply slip contained tick boxes regarding the convenience or otherwise of the interview time, and asked respondents to suggest a different time or place for the interview if not convenient. It did not, though, contain a tick box declining consent to my visit. In my initial letter contact, then, I chose to deceive by omission regarding the principle of informed consent. In short, I used the implied authority of the Probation Service to coerce as many people as possible to see me to discuss the evaluation.

However, after due consideration, and after consultation with a trusted colleague and with my manager, I felt ethically justified in using such coercion given my 'pilot' experience of people not

presenting for office-based interviews when my appointment letters specifically indicated that the interviews were not compulsory. There seemed little chance of causing 'harm' to the recipients of my letter, and whatever the ethical price of proceeding this way, this had to be balanced against the ethical price of not getting to see sufficient numbers of 'hard to reach' INTEGRA clients to statistically validate the evaluation, thereby undermining it, and undermining the 'good', in terms of practice and policy development, that I believed could come from its use. Furthermore, any subsequent interview would still only proceed with the respondent's signed consent.

Of course, obtaining a signature on a form does not necessarily signify *voluntary* and *informed* consent. Properly understood, obtaining consent is a process of communication between the evaluator and the respondent. This places an obligation on the part of the evaluator to ensure that the respondent understands the nature of the evaluation, and what their participation would mean. This does not simply entail ensuring that they can, and do, read an Information Sheet, but also means that the general attitude of the evaluator should be conducive to respondents' understanding and sense of autonomy:

An individual's understanding of the consent statement and acceptance of his or her status as an autonomous decision maker will be most powerfully influenced not by what the individual is told, but by how he or she is engaged in the communication. (Sieber, 1998, p.131)

Given my use of implied Probation Service authority to get to see potential respondents in the first place, I was particularly keen to convey a non-authoritarian and non-judgemental stance in my face-to-face contacts. This included an apology for the 'pushiness' of my initial approach, the rationale for which was explained, as was the rationale for the evaluation itself, including the proposed content of the interview and the individual questionnaires. Taking time to ensure understanding before consent was especially useful, too, in helping to gauge respondents' cognitive functioning given that all were already identified alcohol and/or other drug users. I also asked systematically if people could read the consent form when I handed it to them, and tried to do this in as sensitive a way as possible: 'I hope you don't think I'm being cheeky or anything, but quite a few people I've seen have had problems reading this form. Is it a problem for you too?' And for several people for whom it was, I read the form to them, slowly and word-for-word, and checked that they

understood. I also guaranteed confidentiality, not only for their decision to participate or otherwise, their capacity to read or otherwise, but also for the content of their responses. Here, though, and once again, it is possible to frame an argument that I behaved unethically. There is no such thing as absolute confidentiality (Robinson, 1991), particularly when dealing with illicit drug issues (Fitzgerald and Hamilton, 1996), and with offending issues generally.

By asking people to convey information about their drug use with the Maudsley Addiction Profile (Marsden *et al.*, 1998), and, particularly at T2, by asking about their current benefit fraud and other offending, I was placing them at possible risk if exposed. Whilst I actively pursued my obligation to protect their right to privacy by removing identification from questionnaires and transcripts, by locking identifiers in a separate location, and by erasing the cassettes recording aspects of T2 interviews, I did not explain the legal obligations that could be placed on me to divulge information under certain circumstances. And I cannot claim ignorance of these matters.

Prior to undertaking interviews, I posted an email to a Drug Misuse Research Discussion Forum asking for information relating to confidentiality. A number of responses pointed me to a paper by Fitzgerald and Hamilton (1996). *The Consequences of Knowing: Ethical and Legal Liabilities in Illicit Drug Research*, documents the ethical difficulties, the legal obligations, and the lack of support from their University, experienced by an Australian research team when they were approached by undercover police officers for an 'exchange of information'. Whilst the legal circumstances outlined are specific to Australia, it seems that the UK drug research community sees its position as very similar. That said, I am not aware of any such problems having been experienced thus far by UK researchers. So, what is the legal situation as I understand it?

Drawing on the analysis of Rice and Thomas (2001), I conclude that:

whilst the police cannot demand to see files designated as 'excluded material' by the Police and Criminal Evidence Act 1984 ss 11-13 (PACE), evaluators and other researchers do not generally enjoy such legal protection. In any event, the police can still apply to a High Court for a Production Order on 'excluded material' (PACE 1984 s1), and have used this power in the past to demand journalistic and medical material.

The Crime and Disorder Act 1998 s115 (CDA) is intended to make it easier for information to flow between agencies who have to implement the Act. These 'relevant authorities' as defined for the purposes of the Act are: police, local authorities, health authorities, and the Probation Service (CDA 1998 s115(2)). Universities, as the employers of many external evaluators, are not mentioned, but the Act does refer to: 'person(s) acting on behalf of such an authority' (CDA 1998 s115(1)(g)), which presumably could embrace external evaluators, not least if one of the 'relevant authorities' is paying a sum of money for their contracted service. The CDA 1998 s115 puts no obligations on agencies to disclose information, but reassures those who may wish to do so that they would not be acting improperly. This reassurance counters aspects of the Data Protection Act 1998 (DPA) which holds the principle that information collected for one purpose should not be used for another purpose (DPA 1998 Sch.1, Part 1, para 2). That said, the DPA 1998 also contains its own exemptions relating to crime detection and reduction (DPA 1998 Sch. 2). In many ways, these exemptions complement the CDA 1998 s115, and a joint statement to this effect was issued by the Home Office and the Data Protection Registrar (DPR) in 1998 (Home Office/DPR, 1998).

More recently, the Human Rights Act 1998 (HRA) (implemented 2nd October, 2000) aims to codify privacy protection under Article 8, but exemptions apply if legal arrangements are in place to permit this. The CDA 1998 and the DPA 1998 provide these legal arrangements.

In addition to such legal arrangements which may be seen as particularly relevant to the INTEGRA evaluation, there are other legal limits to guarantees of confidentiality arising from child abuse and neglect, and from threats of harm to oneself and others. Is it ethical, then, for me to have 'promised' confidentiality when obtaining respondents' consent to participate in the INTEGRA evaluation? On balance, I believe it is:

- 1). I considered it extremely unlikely that any legal moves would be made by the police to access the content of these interviews. As I stated earlier, I am not aware of any such problems having been experienced thus far by UK researchers. As for the difficulties described by Fitzgerald and Hamilton (1996), these arose from an ethnography, and a role which placed the researcher in close proximity to illegal activities. Hence the police interest in possible good intelligence. Given such high improbability of harm to any

respondent by such police interest in the INTEGRA evaluation, I felt it ethical not to risk their participation, and the possible negative impact on the evaluation, with unnecessary anxiety.

2). In the highly unlikely event of legal moves to access the evaluation Project data, I was also clear that the possibility of minor legal harm to me for refusing to co-operate by, say, 'hiding' the identifiers, was a price worth paying to protect the right to privacy, whether or not I received backing from my Department or the University. My ethical reasoning was: the public need for evidence-based practice and policy to better understand and tackle illegal behaviour outweighs the public need to prosecute *bona fide* researchers that do not comply with court evidence procedures. I would also resist any consequent risk to the capacity for evidence-based work which could be caused by undermining respondents' trust in guarantees of confidentiality.

3). Should any respondent have caused concern for their, or other's (including children's) welfare, I designed a Consent to Disclose Form. Ideally, I would be able to negotiate respondent's consent for my disclosure of these concerns to a named individual or organisation. Had I not been able to negotiate consent for disclosure, I would consult with senior staff in my department regarding justified infringement of the right to privacy. In the event, this happened on one occasion following the suicide of the respondent as described below:

DISCLOSURE AFTER DEATH

The Maudsley Addiction Profile (Marsden *et al.*, 1998), one of the questionnaires completed with respondents, measures problems across four domains: drug use, health risk behaviour, physical and psychological health, and personal/social functioning. The psychological health domain embraces questions exploring depression and suicide risk. One respondent caused me particular concern for his welfare across this domain. However, he was in regular contact with his Probation Officer and INTEGRA worker, both of whom were aware of his psychological health and previous suicide attempts, and the battles he was currently having with a Health Trust over their refusal to offer him an inpatient detox for heroin addiction. He saw no point in my disclosing to either worker something they already knew. Indeed, he was much more concerned about explaining his detox grievances to me, and

outlining what he and his Probation Officer were planning in order to fight the case. He was scheduled to see his Probation Officer the next day. This man committed suicide a few months later.

Shortly following his death, I received a request from his INTEGRA worker to disclose aspects of my interview with the man, particularly his detox grievances. The INTEGRA worker explained that the Chief Probation Officer was planning to take up the man's case with a Health Trust contact in an attempt to minimise the likelihood of similar occurrences in the future, and that the content of my interview with him may add weight. I explained that I would need to discuss with senior colleagues before I could infringe the man's right to privacy. My own view was that there would be no harm caused to the man and that some good could come from infringing his confidentiality. The only potential harm, it seemed, could be to relatives who may not be aware of his plight, or who may be embarrassed, either for him or for themselves, by any public disclosure of his circumstances. That said, however, his only surviving relative was a sister, who was also aware of his lifestyle and difficulties. Indeed, she had retained close contact with her brother, visiting regularly, and had discovered his body on one such visit. If necessary, I was prepared to visit the sister to discuss her feelings about my disclosing information which her brother had provided to me in confidence.

In the event, I was directed not to disclose by the senior member of staff. If confidentiality had been guaranteed, it could not be broken. I have to say, I disagreed. Apart from the justification to disclose I can see in the ethical principles of 'doing good' and 'avoiding harm', I also feel justification to disclose can be found in the theoretical perspective of 'social justice', so long as, that is, an 'ethics of care' commitment does not neglect consideration of the possible impact on the man's sister. In the event, however, the Probation Service decided not to pursue the matter with the Health Trust, and my ethical dilemma disappeared.

COERCION AS INCENTIVE

As in other studies (e.g. Hughes, 1999), a constant feature of this evaluation has been people failing to keep appointments for pre-arranged interviews. Such non-appearance might constitute an indirect refusal to participate. But, equally, it might also signify that set interview times do not fit comfortably with aspects of a drug user's lifestyle.

Clearly, as outlined in Chapter 4, drug-using probationers pose a significant conundrum for suitable evaluation design – not least, in the absence of a convenient and user-friendly 'gathering' point, and not least in the absence of tangible incentive. As described in Chapter 3, INTEGRA operated throughout County Durham, working one-to-one, and principally in Probation Service offices. There was no 'gathering point'. As for tangible incentive, unlike other work undertaken by our department, this design did not include payment for participation. For the evaluation manager, this was attributed to 'oversight'. For the Probation Service, when asked about building in payment after my arrival, the main issue, I feel, was one of 'political sensitivity', although it can also be seen as embracing an act utilitarian ethical perspective too. How would 'the public' feel if they heard that offenders were being paid?

Leaving aside, until the next section, ethical deliberation regarding the payment of research respondents, without some form of incentive just how do you get to see sufficient numbers of 'hard to reach' respondents in order to satisfy the statistical demands of a quantitative evaluation design?

At a meeting of the Evaluation Management Group on 12/4/99 it was suggested by the Assistant Chief Probation Officer Chair that 'research' contacts be made 'reporting' contacts under National Standards (Home Office, 2000). Failure to keep appointments would be recorded as an unexplained absence. Two such absences usually result in a return to court for Breach proceedings, a likely further sentence, and possible imprisonment. In this way, non-attendance for an evaluation contact was to become 'punishable'. The 'ethical space', I feel, was being re-negotiated. Although this was suggested by way of offering incentive to probationers, offering a 'carrot' (i.e. seeing me would replace their Probation Officer contact that week), the threat of 'stick' is clear too. My ethical reservations about linking possible punishment to research were dismissed at the meeting by my evaluation manager as 'idealistic' and I was directed to proceed down this route to 'get the numbers'. I decided not to argue the case further with my manager at that meeting. But after very careful consideration and consultation, I later chose not to pursue the notion of 'reporting research contacts'.

My plan of action to arrive at this decision was threefold:

- 1). Consult widely but appropriately on the ethics of turning 'research' contacts into 'reporting' contacts. How much would my intuitively felt ethical discomfort be shared by other

evaluators/researchers? Might other perspectives help my own thinking?;

2). Test how it would 'feel' to undertake a research contact as a reporting contact;

3). Seek an ethical alternative approach to 'getting the numbers'. If successful, this would eliminate the dilemma and diminish the pressure I was feeling myself to be under, as well as the stress of any further disagreement with my manager.

Developing each in turn:

1). EVALTALK is the email discussion forum hosted by the American Evaluation Association, but also used by professional evaluators across the globe. I posted my ethical dilemma to the forum for comments. From a subscriber list of around 2,000 evaluators, I received six responses. Five felt it was clearly unethical to consider turning research contacts into reporting contacts. The sixth, however, held a different view:

Offenders have to jump through many hoops on the road to being free – this is just another.

There were two further responses prompted by the above reply, and both clearly surprised that the evaluation process was somehow being endorsed as a component of the criminal justice process.

In addition to EVALTALK, I also emailed my dilemma to a Probation Forum, and received a personal reply from another researcher juggling with a similar ethical conundrum in relation to sex offenders subject to community supervision: how to get sufficient numbers for a quantitative design. What would be the public response to payments as incentive? Should they, instead, change the methodology to a qualitative approach? Should they attempt to 'double-up' research appointments with reporting times, thereby *indirectly* using implicit Probation authority, and National Standards, to coerce attendance? This researcher shared my ethical discomfort at the thought of *directly* linking research contacts to National Standards and potential punishment.

Despite a number of emails and telephone messages, the Home Office Research Unit failed to respond to my requests for advice.

2). In testing the process of linking evaluation contact to National Standards, my baseline principle was to 'avoid harm'. If possible, I also wanted to 'do good'. I discussed the situation with a Probation Officer who was 'reluctantly' about to breach a probationer in my INTEGRA sample. This young man, apparently, was 'doing well'. He had stopped using drugs and was holding down a steady job. His mother, with whom he still lived, had been extremely supportive despite years of distress his behaviour had caused her. He, though, was resentful and angry at having to report to Probation, not least given the changes he had made in his life, and he was now failing to keep appointments. Recent telephone conversations between the Probation Officer and the young man had not been productive. His mother was distressed too, not understanding either her son's attitude, or the Probation response in being prepared to Breach despite the positive changes in her son's life.

With the agreement of his Probation Officer, I wrote to this young man and advised that meeting me to discuss the evaluation would be recorded as an official Probation contact.

In the event, the young man was not at home when I called. But his mother was, and I spent around one hour allowing her to exhale frustration and tears. I was also able to help her understand the Probation stance more clearly i.e. that increasingly Probation Officers have no discretion regarding Breach: the Home Office lay down the standards, and if these are not adhered to, then Probation Officers are held to account. Between us, we arrived at a suitable time for me to call again when her son would be at home – 8pm one evening, one hour before he left home for the nightshift. She agreed to tell her son of my visit, and felt that now she understood the Probation situation more clearly she might be able to help him feel less angry – he too was blaming the Probation Officer for the situation. I chose potentially to compromise my professional standing and ignore National Standards by not informing the Probation Officer of his failure to report for this interview.

The young man was at home for the next appointment, although initially very suspicious. Nevertheless, he consented to the evaluation interview itself. However, it became clear that he did not really want to do it. When I put this to him after ten minutes or so, he readily agreed, but thought he 'had to do it because it was part of Probation'. I tore up the consent form, and told him I would let the Probation Officer know he had 'reported', but that what had transpired was confidential and the officer would not find out whether we had completed the interview or not. I encouraged him to

contact the officer, though, to sort out his appointments, and that it would be a shame to see his hard work 'knocked back' by a court appearance for a 'silly' Breach, and then I headed for the living room door. At which point his mother came in, surprised then tearful and upset because my leaving so early meant her son had not completed the interview, and concerned that this would lead to his being Breached. She too, despite my assurances otherwise, had also thought he 'had to do it because it was part of Probation'. Fifteen minutes later, I was able finally to leave with calm restored, and with a sense of having done some 'good'. But I resolved not to accept the Probation Service offer of making evaluation contacts into 'reporting contacts'. I had to find an alternative.

3). In the event, there was no single magic alternative, simply a concerted effort to refocus some of what I was already doing, and a recognition that energy and tenacity were my best friends. A meeting with a group of probationers (see Chapter 4) helped shed some light, particularly regarding the value of home visits, and the lack of value and respect in 'cold calling'. And whilst all of this resulted in long days on the road; numerous repeat visits to missed appointments; various chats with informal gatekeepers (mothers, lovers, and others) who would tell 'what I was like' and pass on messages; and hours parked up on County Durham highways and byways waiting for the next appointment time, it all came with a sense of relief that an increase in numbers had resolved my ethical dilemma without any risk of 'harm' to the INTEGRA clients.

In the event, 321 home visits resulted in 75 successful interviews at T1, around 10% of all those seen by the Project. But would my ethical choice be regarded in the same light had I interviewed many fewer respondents, thereby, presumably, undermining the principle of 'doing good' that could come from the work, and, possibly, impeding the principle of 'faithful' fulfilment of an agreed contract? Were I a salaried researcher, would cost then prohibit the luxury of necessary perseverance? In such circumstances, would re-negotiating the methodology and the contract, whatever the inconvenience, be more ethical than redefining the boundaries of the ethical space itself? Notwithstanding the substance of such questions, there is, perhaps, an even more salient one....

Is this quandary something other than ethical?

On 2nd October 2000, the Human Rights Act (HRA) 1998 integrated the European Convention for the Protection of Human Rights into UK law. All public authorities, and their contracted agents, are now subject to the scrutiny of human rights, and subject to possible legal

challenge. Without doubt, test cases can be expected to intensify, and the Act brings lawyers ever closer to the practice and policy-making agenda.

With the INTEGRA evaluation in mind, where would the notion of making any research contact, for whatever reason, 'enforceable' stand in relation to the Act? Certainly, it seems questionable under Article 8, the right to privacy in private and family life. Likewise, had any failure to attend for a research contact contributed to a probationer being imprisoned, then challenges under Article 5, the right to liberty, and perhaps Article 7, freedom from arbitrary punishment, may not be unwarranted? But this is speculation. I am not a lawyer and no such test cases have been reported.

There is one further ethical issue which the dilemma regarding the use of National Standards to encourage respondent participation raises:

The ethics of being user-focused

Beyond general ethical sensitivity and awareness, the adoption of a utilisation-focused approach, as with the INTEGRA evaluation, and where the aims are to limit stakeholder involvement to primary intended users and then work closely with these, raises specific ethical concerns that require some comment.

Such issues have already been touched upon in Chapter 2, with Pawson and Tilley's (1997) critique of Patton. Whilst I believe Pawson and Tilley were unfair with aspects of their stance, the issue of 'who owns the evaluation process' is clearly an important one. As Patton (1997) himself recognises:

Concern that utilization-focused evaluators may be co-opted by stakeholders, or become pawns in the service of their political agendas, raises questions beyond how to be politically astute...or prevent misuse...decisions about one's relationships with intended users involve ethics. (p.362)

Not only is power involved here, but money is involved too. 'He who pays the piper might not always call the tune', but such issues demand alert ethical awareness on the part of any evaluator professionally committed to honest reporting and yet personally reliant on monetary gain or future work.

But the dangers of co-option are not necessarily driven by power or money. They may involve something much more subtle, but

something just as risky – the inculcation of an organisational culture which could prove the downfall of any evaluation's integrity. Even though I believe the Probation Service offer to make evaluation contacts 'reporting contacts' was made from a genuine commitment to be helpful, had I accepted and proceeded down this route I believe I would have committed one of the most serious ethical errors of utilisation-focused evaluation. I would have allowed myself to become co-opted by the cultural alignment of the primary intended user, to become co-opted by the enforcement and control imperative of the Probation Service, rather than retaining my independent standing as an external evaluator.

PAYMENTS FOR PARTICIPATION

As mentioned earlier, I was unable to offer tangible incentive at T1 to encourage participation. In large part, this was due to Probation Service concern for public opinion regarding payments to offenders. However, having presented CDPS with a Home Office sample cost schedule for the engagement of external evaluators (in Chapman and Hough, 1998), a schedule which includes an interview payment of £10 per respondent, CDPS felt able to agree such payment for T2 interviews.

This issue of payment for participation, though, begs some ethical comment. Does it, for instance, represent a form of coercion? Is it possible, as Homan (1991) argues, that in relation to the question of informed consent, 'payment for participation constitutes an improper influence' (p.71), or in some way tarnishes the data? For me, though, payments to the INTEGRA respondents were simply a recognition of the fact that they were giving of their time, and not least at Time 2 when the greater qualitative supplement to the interviews required even more effort on their part. That said, the money, whilst seen as something of a bonus, had not been anticipated by the respondents who had already agreed to be interviewed 12 months earlier without any payment. In my opinion, payments, quite simply, are about respect, and represent a commitment to social justice where people receive some recognition for giving their time and thought for our benefit as researchers or evaluators – the same recognition, I am sure, that many would have less difficulty with if our respondents were surgeons or solicitors – although I very much doubt that a rate of £10 per hour would suffice if that were the case.

CONCLUSION

In general terms, upon reflection of my stance in relation to the ethical conundrums outlined here, I feel there is a strong act (consequentialist) utilitarian thrust to my ethical thinking. However, as evidenced in my attitudes towards 'reporting research contacts' and 'payments for participation', there is also a strong social justice element, along with an ethic of care, and a commitment to advocate for the needs of the disadvantaged. Perhaps this is not surprising given my social work occupational choice.

The Guiding Principles of the American Evaluation Association, taken together with Newman and Brown's (1996) framework for ethical decision-making, make it clear that evaluators, from the very start of an evaluation to the very end, encounter various situations which demand a grounding in applied ethical thinking. This chapter aims to show how this goes far beyond the, at best, ethical minimalism of a reliance on codes or the 'outset approval' of some Ethics Committee or other.

However, this chapter does not attempt to argue that my navigation of the ethical space within the INTEGRA evaluation, somehow, 'got it right'. Being ethical is much more than making, what we might consider to be, sound ethical choices regarding specific incidents or situations in one specific evaluation; being ethical is a professional way of life, and like any way of life this means we must be constantly alert to any tendency towards self-righteous complacency, and humbly alert to the realisation that we will not always make the best decisions or the best choices. But being ethical does mean that we will keep trying.

Chapter 6

ILLUMINATING THE FINDINGS

Book-burning fire captain: Chock them so damned full of 'facts' they feel stuffed, but absolutely 'brilliant' with information. Then they'll feel they're thinking, they'll get a sense of motion without moving.

Ray Bradbury, *Fahrenheit 451*, 1953, in Joel L. Swerdlow, 'Information Revolution,' *National Geographic*, October 1995

INTRODUCTION

This chapter presents the main findings from the INTEGRA evaluation as detailed in the Final Evaluation Report. The data were all obtained by me, as described in Chapters 4 and 5. The findings from my investigation of primary intended user response to the Final Evaluation Report, and the evaluation itself, are covered separately in the next chapter.

At the outset, it is worth repeating points made in Chapter 4 regarding data analysis. My involvement in the analysis of core quantitative data was through checking data entry on FORMIC, editing, and the production of descriptive statistics. More detailed inferential statistical analyses were undertaken by the evaluation manager, Dr. Justine Schneider, and by Dr. Pauline Coolen-Schrijner from the Statistics Consultancy Unit at the University of Durham. Analysis of all qualitative data, along with the 'Motivation to Participate' and 'Case Manager Survey' components was undertaken by myself.

The presentation and interpretation of the data is my own and does not necessarily accord with the interpretation of the evaluation manager. That said, the basic interpretation and presentation does not differ markedly from that which was negotiated between the manager and myself for the Final Evaluation Report (Burlison and Schneider, 2002). As such, what is presented in this chapter is in essence what was presented in

the Final Report, and what the primary intended users of the evaluation are responding to in their reading of this Report outlined in the next chapter. For this thesis however, I have included considerably more qualitative data reporting the views of INTEGRA users and staff. I have also extended the interpretive comments in order to enable me to reflect, in the Chapter 8 Discussion, on the theory and research presented in Chapters 2 and 3.

DESCRIPTION OF THE CLIENTS

THE SAMPLE

Included in the evaluation sample of INTEGRA clients were all those people who had been referred and seen by an E Team Project worker for mentoring between June and December, 1998 (Cohort 1 at Time 1 [T1]), and between June and December, 1999 (Cohort 2 at T1). Although the two groups were all referred to INTEGRA, Cohort 1 were more likely to be individuals who had been on Probation Officers' caseloads for some time, presenting problems with alcohol and/or other drugs, whereas Cohort 2 were more likely to be referred to INTEGRA at the same time that they were given a Probation Order.

Table 6.1 below summarises the numbers involved in the sample, and the numbers successfully interviewed for both cohorts. Table 6.2 summarises the characteristics of the sample. Table 6.3 outlines the numbers from both cohorts who were successfully re-interviewed for the study 12 months later at Time 2 (T2).

Table. 6.1: Numbers of respondents at T1

	COHORT 1	COHORT 2
Referrals	79	108
Adjusted*	56	83
Refusals to participate	7	14
Uncontactable	10	19
Possibly contactable	0	14
Respondents	39	36
Number of appointments	146	175
Ratio of appointments to interview	3.7	4.9

*when adjusted for people who had left area/ no forwarding address/ in prison in another part of the country/ warrants out for their arrest/ in breach of Order/ deemed 'unsuitable for research contact' by Case Manager. Regarding this last adjustment, only one person (in Cohort 1) was deemed 'unsuitable' by their Case Manager given their state of mental ill health.

Table 6.2: Sample characteristics at T1

Characteristics of sample	Cohort 1	Cohort 2
Male	95%	86%
Mean age at referral (Std. Dev.)	26 (8)	29 (10)
	Range 18-49	Range 17-49
White British	100%	100%
Education beyond age 16	21%	22%
Living arrangements:		
Alone	23%	20%
With partner	33%	42%
With parents	36%	28%
With relatives	5%	6%
With others	3%	6%

Table 6.3: Numbers of respondents at T2

	N
T1 Respondents (Cohort 1 = 39 + Cohort 2 = 36)	75
Refusals*	7
Deceased	1
In prison in another part of the country	1
Not found and not current on Probation Service records**	21
Total of T2 respondents	45

*the people who refused had finished their Orders and were adamant that they wanted "nothing more to do with the Probation Service".

**the numbers of people not found, having left their contact address from T1 and not being current on Probation Service records could, possibly, have been reduced had I left a stamped addressed envelope for them to contact me should they change address.

In Cohort 1, six interviews were held in probation offices to 'double-up' with normal 'reporting'. This reflected my use of implicit Case Manager authority with some people who were proving difficult to contact, but were equally possibly 'available' as they were maintaining regular Probation contact. Given that 21% of Cohort 1 said, at the end of their interviews when asked, that their responses to my questions 'would have been different had they been interviewed in a probation office', and given the views of the small group of probationers I met, I decided to keep Case Manager involvement to an absolute minimum for Cohort 2, and only to interview in a probation office when this was specifically requested by the client, or whenever it was indicated on health and safety grounds. In the event, no Cohort 2 interviews were undertaken in an office.

The total number of interviews undertaken for both cohorts, n=75, represents around 10% of all clients seen by the E Team.

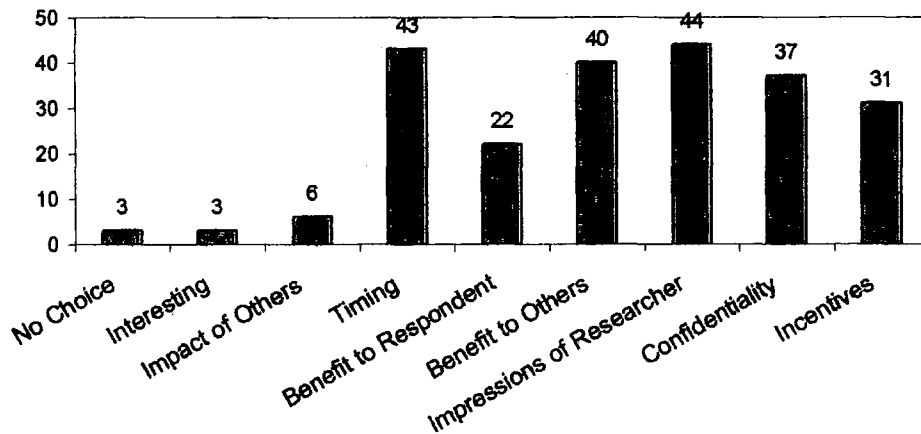
What is worth noting, at this point, is my abortive attempt to construct an 'experimental' control by repeating this cohort approach with a group of 'standard probationers' randomly sampled from the CDPS database during the time period corresponding to Cohort 1. In the event, three weeks of solid attempt resulted in six completed interviews. It also resulted in more doorstep antagonism to my contact. This would seem to validate the point of Hughes (1999), and the views of the probationer group, that the research should have some relevance for the respondents. Generally, people who had "never heard of INTEGRA" were much less willing to be involved.

T2 interviews were accompanied by a payment of £10, in cash at interview end, and with a receipt signature. The issue regarding interview payments has already been covered in Chapters 4 and 5. However it is worth repeating that it arose from T1 concern regarding the numbers of people not keeping to pre-arranged appointment times. At T2 this issue prompted me to introduce some systematic investigation of people's motivation to participate in the evaluation. This might help in exploring the possibility of 'contactability bias'. Perhaps the evaluation recruited the more co-operative service users?

MOTIVATION TO PARTICIPATE

As outlined in Chapter 4, few studies seem to address the question of people's motives for participating in social research generally. What is reported tends to be the reflections of the researchers. As Lee (1995) points out, what is lost are the perspectives of the participants. In his qualitative study of 17 drug injectors, Hughes (1999) reports a variety of reasons for participation, ranging from altruistic motives to instrumental ones. Figure 6.1 summarises the explanations for participation given by the 45 respondents successfully re-interviewed at T2 and based on Hughes' themes.:

Fig. 6.1: Reasons for Participation



Despite my efforts to ensure that people understood they did not have to participate in this evaluation, three people felt they “had no choice”. The most commonly reported reasons for participation are:

- **The timing** – people appeared very positive about home visits and the inclusion of stamped addressed envelopes and reply slips in contact letters so that they could re-arrange the appointments if they wished. As in Hughes’ study, a feature of this evaluation was people failing to keep appointments for pre-arranged interviews. Such non-appearance might constitute an indirect refusal to participate. But it might also signify that set interview times, particularly where some organisation and effort is required to get to an office, do not fit comfortably with some aspects of a drug user’s lifestyle. Rather than interviews in an office, then, home visits appear to be more successful.
- **Impressions of the researcher** – bearing in mind that people may have felt pressure to give a desirable response in my presence, people appear to have responded well to my approach and style.
- **Benefit to others** – 40 people felt that their participation might have some benefit for others. One would need to question such altruism given the social desirability of such responses.
- **Confidentiality** – perhaps not surprisingly given some of the ‘sensitive’ content of the interviews, 37 people felt that my

reassurance regarding confidentiality was an important aspect of their agreement to participate.

- **Incentive** – at T2, participation was encouraged with a £10 incentive. While 31 people felt that this was a useful way to encourage participation, many people qualified their views with a “It’s not that important for me though, but I can imagine it would be for others. But I saw you last year anyway and I said then I would see you again.” The real benefit of payment for this study may have been seen at T1 when people had no experience of myself or the study.
- **Benefit to respondent** – 22 people felt that the interviews may have some benefit for them by ‘talking about things’. This echoes the question of contactability bias throughout this study: perhaps the evaluation recruited those people who were more keen to do something about their situation? Interestingly, in this vein, 6 people felt that ‘the impact of others’ was a factor in their decision to participate. The common theme in their responses was the impact of their INTEGRA worker, the fact that the respondent had experienced the Project positively and “wanted to give something back”.

Motivation for research involvement, then, at least for the 45 respondents re-interviewed for this study at T2, appears to be multi-dimensional, not motivated by economic gain alone, and not necessarily defined by any direct benefit or gains to themselves. This is consistent with the findings of a recent Australian study involving 154 injecting drug users recruited through needle and syringe exchange programmes (Fry and Dwyer, 2001), and does not provide evidence of ‘contactability bias’.

‘CONTACTABILITY BIAS’

Since INTEGRA was an innovative programme, targeted at offenders with alcohol and/or other drug problems, it was important to discover how its beneficiaries might differ from other people on Probation. The Level of Service Inventory-revised (LSI-r) is a useful tool for such comparisons, because it incorporates many dimensions related to risk and social problems (Raynor *et al*, 2000). It was also administered to a large proportion of CDPS clients during the research period. Nevertheless, the data available for making comparisons were incomplete because in several cases LSI-r ratings were missing. For Cohort 1, 28 out of 39 people, and, for Cohort 2, 25 out of 36 people had LSI-r data.

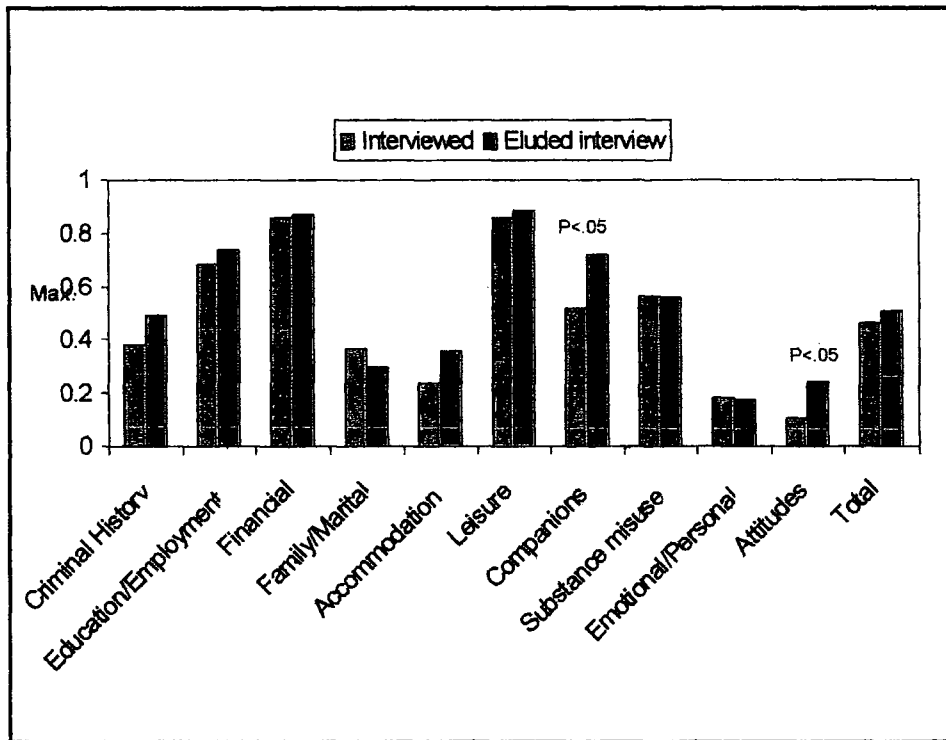
LSI-r data also allowed exploration of differences between those interviewed for the valuation and those not. Table 6.4 shows that the people who were interviewed were less likely to have several characteristics. These reflect first, a history of offending, secondly insecure housing (which could be directly related to failure to contact) and thirdly a social network that contains acquaintances and friends who support criminal activity.

Table 6.4: Differences between respondents and non-respondents (Fisher's Exact test)

Clients interviewed less likely than those not interviewed:	Level of significance (p)
To have two or more prior convictions	.041
To have been arrested under age 16	.040
To live in unsatisfactory accommodation	.008
To have few anti-criminal acquaintances	.006
To have few anti-criminal friends	.008

The 54 questions of the LSI-r variables can be combined into a number of sub-scales. Figure 6.2 illustrates the differences between the clients interviewed and those not interviewed in relation to these scales. Only two of these scales showed a difference between the two groups that achieved a level of significance that was less than 5% likely to be due to chance. The scale that measures a predominance of criminal companions indicates that people interviewed were less inclined towards such company. The components of this scale are shown in the bottom two rows of Table 6.4. On the second scale where differences were found, which is an indicator of negative attitudes towards the Probation Service, those people not interviewed scored higher. This is consistent with the idea that some deliberately avoided being interviewed. The data from the LSI-r, therefore, suggest that the people whom I did not succeed in interviewing had more criminal companions and were less positive about receiving help. This is evidence of 'contactability bias'.

Fig. 6.2: Comparisons between those interviewed and those not



I also explored the case files that were available for the people not seen, to see if I could find out more about them. Of the 40 people who were not seen for one reason or another, 32 had assessments recorded on file. Of these, 19 (59%) had problems with drugs other than cannabis or alcohol. Therefore, it seems reasonable to infer that a higher proportion of people not seen were using illicit 'hard' drugs. This is further evidence of 'contactability bias'.

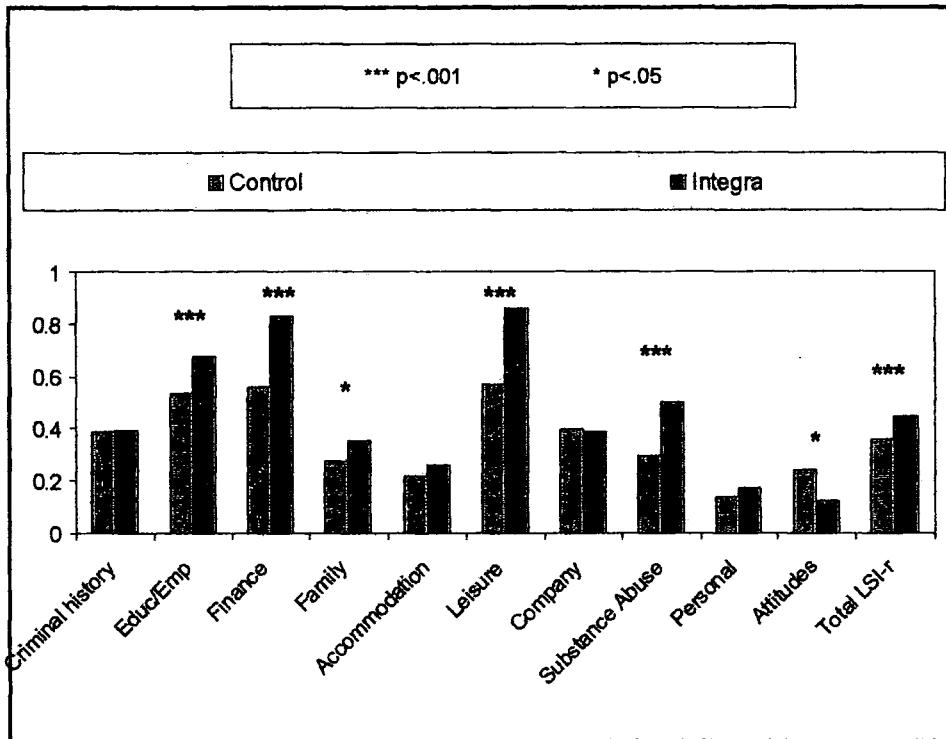
THE SAMPLE COMPARED TO OTHER PEOPLE ON PROBATION

Cohort 2 was compared to Cohort 1 in terms of the LSI-r scales, and was found not to differ on any of these. Therefore all those people interviewed in both cohorts for whom there were LSI-r scores (N=53) were compared to the control group (N=573). The control group consisted of all other people on Probation on the County Durham Probation Service database for whom LSI-r scores were available. Figure 6.3 shows the results of this analysis.

Table 6.5 gives the mean item score on LSI-r scales and the significance of the differences using the non-parametric Mann-

Whitney/Wilcoxon test. All of these results were confirmed by analysis of variance.

Fig. 6.3: Comparison between INTEGRA clients and controls on LSI-r



As shown in Table 6.5 below, the study sample scored significantly differently from the controls on 7 out of 10 LSI-r scales. On six of these, including the LSI-r total, they scored higher than controls. On attitudes towards crime and sentencing they scored lower. These results indicate that the INTEGRA clients had more problems than the norm for Probation clients, and these problems were in relation to education and employment, finance, family, leisure and, not surprisingly, substance use. Clearly, alcohol and/or other drug problems may lie at the root of any or all of these problems.

By contrast, the study sample had more positive attitudes towards crime and sentencing. This is surprising in view of the fact that the Project was designed so that there should be no filtering of candidates prior to referral, but that this should be automatic where the LSI-r of a person on a Probation Order showed problematic drug or alcohol use. However, in the early months of the programme,

Case Managers were free to refer their existing clients, and anecdotal evidence indicates that there was some selection. Some E Team staff in their exit interviews, as described later, believed this selection tended towards more difficult clients and there was, for them, a sense of being “dumped on” by Case Managers in the early months of the Project. However, the evidence here may suggest that some Case Managers may have referred existing clients who were more likely to benefit from the new Project.

Table 6.5: Differences in LSI-r subscale (item mean) between control group and INTEGRA clients (Wilcoxon signed-ranks test)

Level of Service Inventory Sub-scale	Controls mean N=573	INTEGRA mean N=53	Z	Significance
Criminal history	0.39	0.39	-0.14	0.8891
Education & Employment	0.53	0.68	-3.39	0.0007
Finance	0.56	0.83	-4.87	0.0000
Family	0.27	0.35	-2.25	0.0244
Accommodation	0.22	0.26	-1.51	0.1313
Leisure	0.57	0.86	-4.67	0.0000
Companions	0.40	0.39	-0.42	0.6724
Substance Abuse	0.29	0.50	-5.35	0.0000
Personal	0.14	0.17	-1.80	0.0717
Attitudes	0.24	0.12	-2.29	0.0223
Total LSI-r	0.36	0.45	-3.65	0.0003

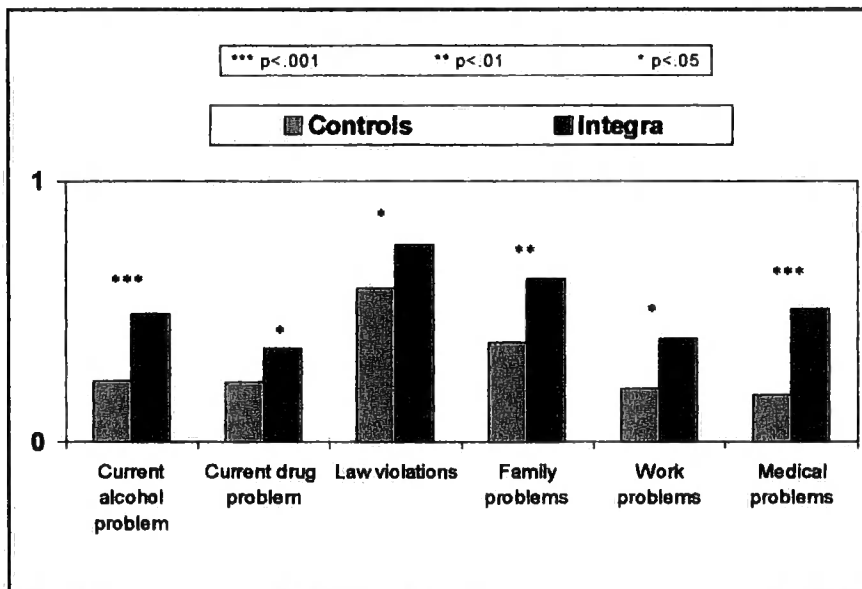
Table 6.6 gives details of the components of the employment and education subscale, showing that the determining factor of the difference between the two groups was employment, rather than education. In terms of employment status, then, INTEGRA clients were more likely than controls to be out of work. 20% of them were working as compared to 34% of controls ($p=.003$). There were no significant differences between the two study cohorts with respect to any of these variables. It is important to note that, while the INTEGRA sample did have higher levels of current substance use, they were no worse off than other CDPS clients in terms of education.

Table 6.6: Frequency of education and employment-related difficulties in control group and INTEGRA clients interviewed (Chi-Square test)

	Control group	INTEGRA	X ²	p
Currently unemployed	67%	94%	17.47	.000
Frequently unemployed	53%	70%	5.35	0.014
Never employed for a full year	37%	40%	0.118	ns
Ever dismissed	20%	30%	2.86	ns
Left school at 16	82%	79%	0.228	ns
Left school without qualifications	63%	55%	1.17	ns
Suspended/expelled	25%	26%	0.08	ns

Figure 6.4 shows all the social problems from the LSI-r on which the study and control groups differed significantly. The INTEGRA clients were also more experienced offenders. Although 16% were not on a current Court Order when interviewed for the study, probably due to the lapse of time between admission to INTEGRA and incorporation in the study sample, half of each cohort had spent time in jail as compared to only 38% of the controls. On average, total time spent in jail came to 24 months (range 1 month to 14 years).

Fig. 6.4: Social problems: Comparison between control and INTEGRA samples (Chi-Square test results)



Interpretation of the differences

Since INTEGRA was designed for people who were unemployed and had problematic drug or alcohol use, it is reassuring that these findings confirm that these factors were significantly higher in the study sample than in other clients on Probation. Thus, it is possible to say with confidence that the INTEGRA Project successfully reached its target group. Overall, the INTEGRA caseload was a more problematic group in relation to the LSI-r and custodial history than the rest of the Probation Service caseload.

WORK STATUS OF STUDY SAMPLE

Taking both cohorts together, 15 clients were working at the time of their first interview. Eight had never worked. Therefore, 52 people were not currently working but had done so in the past. Of this number, 27 were in cohort 1 and 25 in cohort 2. The median time out of work was 24 months (mean 43.7; range 1 – 156) for Cohort 1 and 16 months (mean 37.4; range 1 – 228) for Cohort 2. Therefore, Cohort 2 had been out of work less time than Cohort 1.

OFFENDING HISTORY

It has already been shown above that the people who used INTEGRA had greater exposure to criminal justice services than other people on probation. Table 6.7 shows the offences for which they were convicted, and Table 6.8 describes the mean amount of time they spent in prison. This was much greater for Cohort 1 than for Cohort 2. Again, as discussed earlier in relation to the selection process, this may be a result of the Case Managers selectively referring more difficult clients at the start of the Project, which would have affected Cohort 1 more than Cohort 2. This would validate the views of some E Team staff expressed in their exit interviews that they were “dumped on” by Case Managers in the early months.

Table 6.7: Principal offence category

Offence	Cohort 1 Cohort 2	
	Time 1 N=39 %	Time 1 N=36 %
Acquisitive	35.9	33.3
Violence/order/damage	35.9	25.0
Drugs	15.4	13.9
Alcohol	2.6	2.8
Driving general	7.7	8.3
Driving drink/drugs	0	16.7
Breach	2.6	0

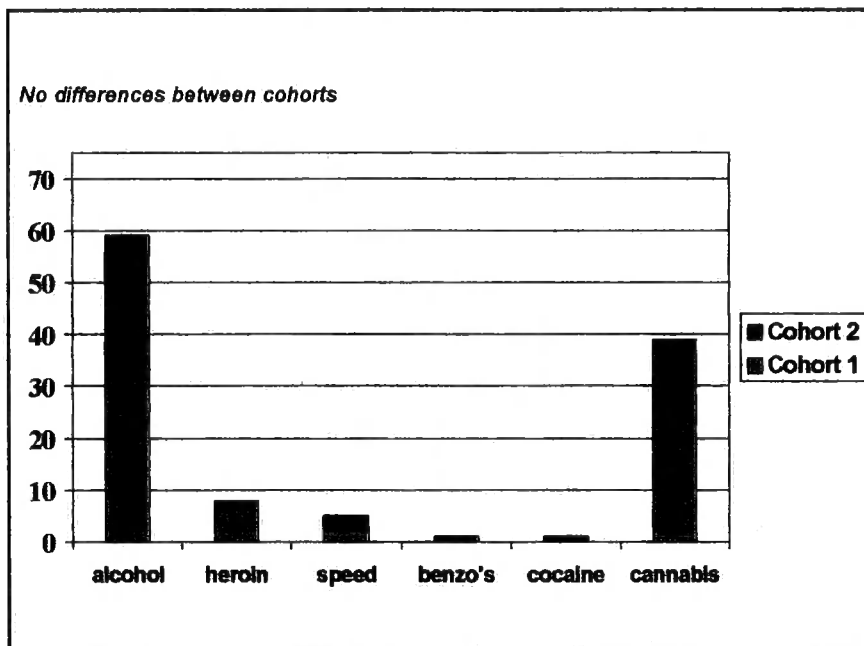
Table 6.8: Total months in prison

	Cohort 1 Cohort 2	
	Time 1 N=39	Time 1 N=36
Mean	9.56	1.92
Standard deviation	27.89	28.61
Minimum	0	0
Maximum	168	140

ILLCIT DRUG USE

Since problem drug use was one of the targets of INTEGRA, we looked in greater detail at the individuals who told me that they used drugs other than alcohol. Figure 6.5 illustrates the distribution of different types of substances used. The total comes to more than 75 because some people had problems with more than one drug.

Fig. 6.5: Problem drug use



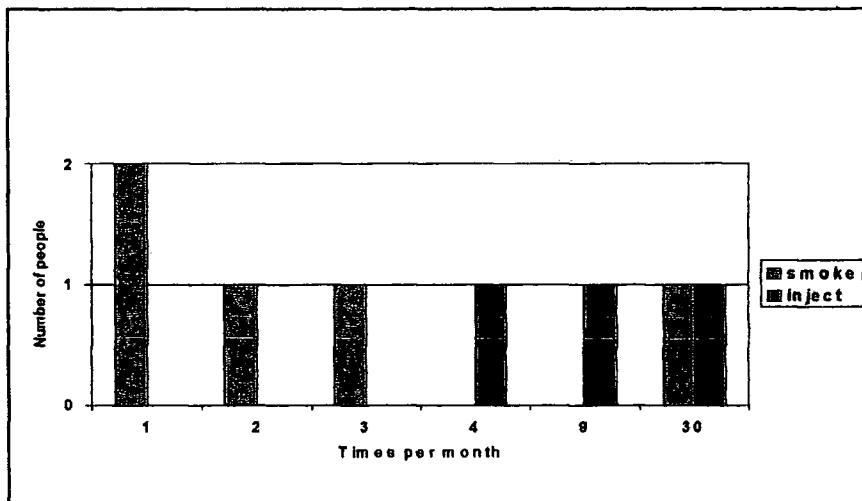
Fewer than 10 of the 75 people in the INTEGRA sample used heroin, five used amphetamine sulphate (speed) one used benzodiazepines (which may have been prescribed) and one used cocaine. The reliability of these figures must be questioned in relation to illegal drugs. Nevertheless, it is clear that the most commonly used drugs were alcohol and cannabis.

Table 6.9 shows the frequency with which people used 'hard' drugs (other than alcohol, cannabis). This is likely to be related to the type of problems they encounter and especially to how difficult they might find it to retain employment. The route of administration is of interest because of the risk of infection for people who inject. Five respondents informed me that they had injected in the past month. Two had done so on one occasion, one on two days, one on eight days and one on 30 days in the past month. Three people said they had injected once per day, one person twice and one once per 'injecting day'. Four people said they had not shared needles (one missing). Taking heroin alone, Figure 6.6 illustrates frequency and route.

Table 6.9: Frequency and route of heroin and amphetamine use

	Frequency Days per month	Route
Heroin (9 people)	Range 1-30 Mean 10 Median 3-4	5 smoke 3 inject
Speed (5 people)	Range 3-9 Mean 6 Median 6	1 oral 2 smoke 1 inject

Fig. 6.6: Frequency of heroin use by means of administration



Another way of measuring drug and alcohol use is in terms of the financial costs to the individual. I was able to estimate this for this study. The costs used are shown in Appendix H. They are based

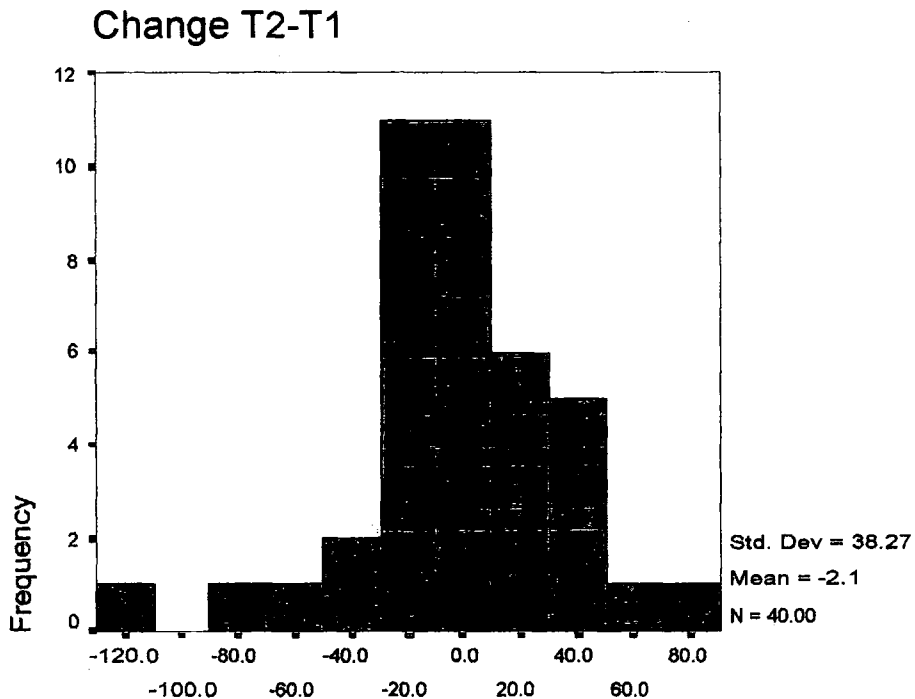
on pub prices, if the interviewee used these, supermarket prices, if they used these, and, for street drugs, Drug Squad information for January 2001 (after the end of the INTEGRA intervention phase).

Table 6.10 shows the minimum, maximum and mean expenditure on drugs and alcohol taken together at T1 and T2, combining both cohorts. The mean change over time for those clients for whom I had data at T1 and T2 (£-2.14) was not statistically significant. Figure 6.7 shows the distribution of changes, from which the change in expenditure on drugs and alcohol can be seen to be roughly equal; 52.5% reduced, and 47.5% increased over time.

Table 6.10 Expenditure on drugs and alcohol

Mean weekly cost (£)	N	Minimum	Maximum	Mean	Std. Deviation
Cost of alcohol and drugs T1	66	.40	424.25	57.15	68.69
Cost of alcohol and drugs T2	42	1.80	185.80	43.87	39.84
Change in cost T2-T1	40	-125.80	77.00	-2.14	38.26

Fig. 6.7: Change in expenditure T2-T1



The fall in expenditure between T1 and T2 can be attributed to random variation, and is not statistically significant for those people whose expenditure is known at both time points.

Implications of the findings

The people in the study spent roughly £50 per week on drugs and alcohol. This did not change significantly over the intervention period. There was no difference in expenditure at T1 between those who were contactable at follow-up and those who were not interviewed a second time.

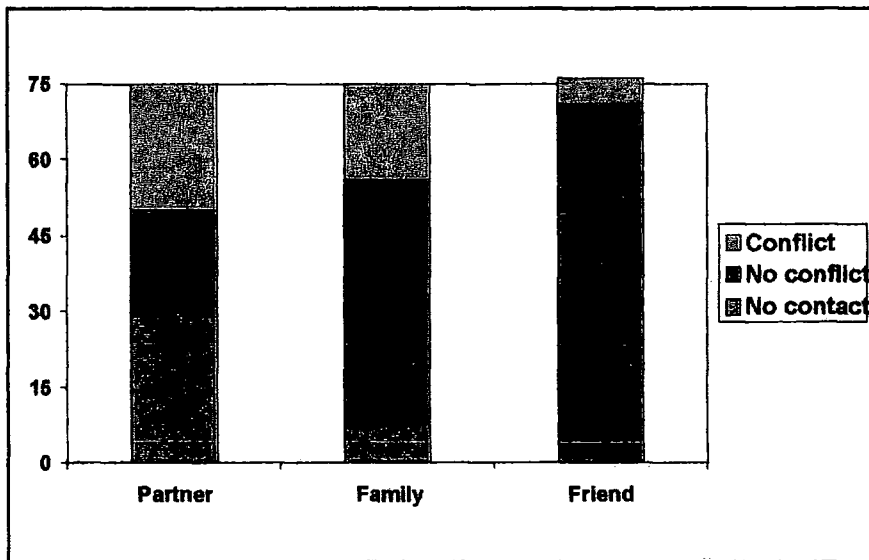
So far, I have explored particular characteristics of the INTEGRA sample interviewed in relation to those not interviewed, and in relation to other Probation clients. These differences are associated with drug and alcohol problems, unemployment, and exposure to criminal justice services. However, other areas of difference were found, notably in relation to what the LSI-r calls family problems and health problems. These may be closely related to alcohol and drug issues.

HEALTH AND SOCIAL PROBLEMS

The Maudsley Addiction Profile (MAP) (Marsden *et al.*, 1998) is a questionnaire designed to explore these difficulties. Whilst there is no MAP data for the controls, the results of this questionnaire help to understand the INTEGRA clients.

The MAP asks respondents whether they have experienced conflict with family, friends and partners in the past month. Figure 6.8 shows that the results differ dramatically according to which relationship applies. Conflict occurred most often with partners; 50% of all those who had contact experienced conflict. Of those people who had contact with family, about one third experienced conflict. However, conflict characterised only about 10% of encounters with friends.

Fig. 6.8: Conflict experienced in past month



Health

The MAP asks detailed questions about health, including both physical and psychological symptoms. Apart from one symptom (muscular pain) which was more frequent for Cohort 1 ($p < .05$), symptoms did not differ between cohorts, as confirmed by Chi-square tests on the testing the presence/absence of each symptom for each cohort. Therefore, in figures 6.9–6.11, the INTEGRA sample is not differentiated by cohort.

In terms of physical symptoms, about half the sample people experienced fatigue often or always, but all other effects were mostly infrequent. In terms of psychological symptoms, tension, panic and lack of interest were the most common, but again, most people only experienced psychological symptoms 'rarely' or 'sometimes'. The age of the client was positively correlated with physical symptoms but not with psychological symptoms. This is consistent with an increase in physical health problems with age in the general population.

Fig. 6.9: Maudsley Addiction Profile symptom inventory I

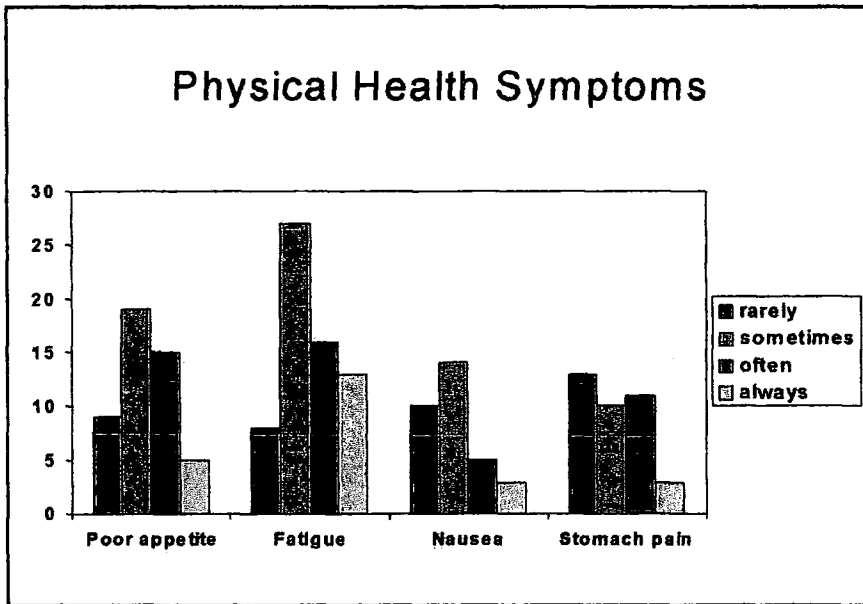


Fig. 6.10: Maudsley Addiction Profile symptom inventory II

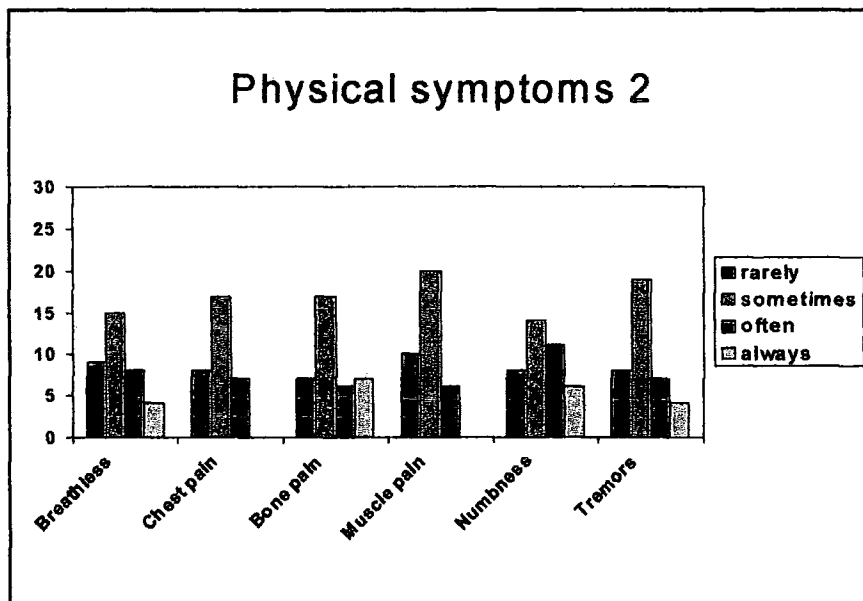
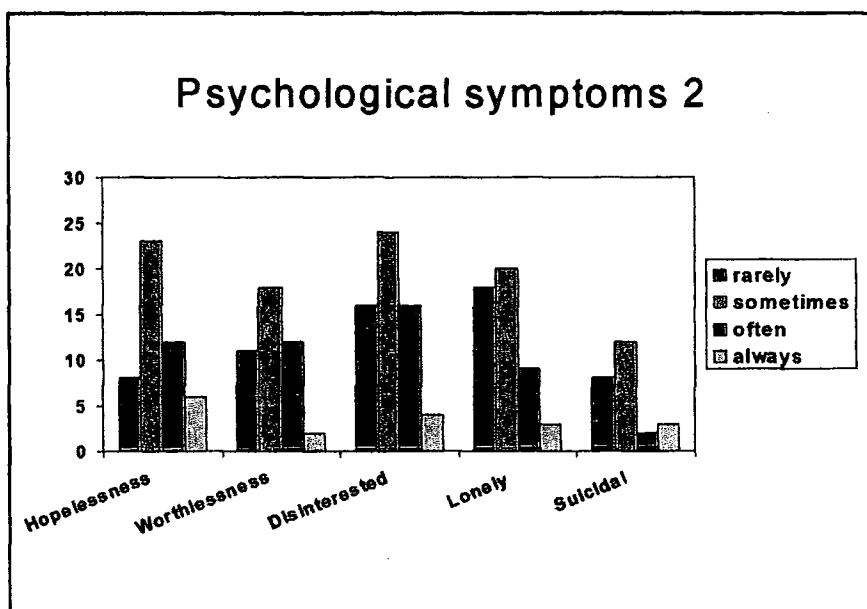


Fig. 6.11: Maudsley Addiction Profile symptom inventory III



Having described the INTEGRA Project clients and having explored their motivation to participate in this evaluation, and before moving on to consider Project Outputs and Outcomes in more detail, I would like to consider the qualitative aspects of their data.

QUALITATIVE RESPONSES FROM CLIENTS

As outlined in Chapter 4, I gathered Project attenders' views at both T1 and T2. Semi-structured interviews were used, with CDC EZ-Text software to facilitate the analysis of the responses gathered. At T1, the interview took the simple form of a Balance Sheet, in the respondent's own words, of the pros and cons of their contact with INTEGRA. At T2 I gave them a longer interview exploring:

- views of the INTEGRA premise i.e. 'unemployed offenders who find a job are less likely to re-offend than those who remain jobless' (also explored with Project staff and in the survey of Case Managers);
- understanding of, and participation in, 'informal' economic activity – an issue which arose in an anecdotal and unsystematic way during T1 interviews;
- the impact of INTEGRA contact on attenders' attitudes to offending and drug use;

- the impact of INTEGRA contact on attenders' work behaviours and attitudes;

In this way, and particularly by developing themes arising from T1 Balance Sheets, the logic of my approach here can be seen to be based in the qualitative research approach of 'grounded theory' (Glaser and Strauss, 1967). This aims to allow the emergence of an *inductively*-derived, empirically-grounded theory, rather than test hypotheses as in *deductive* approaches. It is well suited to exploratory studies which seek to elucidate respondents' own understanding of their experience, since it begins with broad initial research questions, which are then progressively refined and developed during the research process. By so doing, the approach can lead to the development of theory which resonates with respondents' lived experience, and the contextual particulars of their lives. By the same token, "the openness of the grounded theory approach allows a greater latitude for the discovery of the unexpected, some regularity...totally unanticipated by the concepts that might make up a particular theory or hypothesis" (Rubin and Babbie, 2001, p. 392).

Wherever possible, and with the usual assurances of confidentiality, T2 interviews were tape-recorded to avoid the need for detailed note-taking. Of the 9 interviews not taped, only 4 were due to the interviewees' discomfort or anxiety. The remaining 5 were not recorded because of difficulties with the interview environment, mainly noise, and the likelihood that the recording quality would be so poor as to render it useless.

At T1, I assigned Balance Sheet responses to one of a number of categories and constructed a frequency table. On the Balance Sheet, 'good' things about INTEGRA outweighed negative responses by 3 to 1, as shown in Table 6.11. Apart from two people who felt they simply "had to do it – had no choice", and one who could recall no contact, all respondents had something positive to say about their INTEGRA contact. People generally responded positively to the therapeutic stance of the Project. It was said to give hope or motivation by 'helping' people to 'think positively', it fostered trust and self-esteem by listening and being non-judgmental, and it enabled people to talk about their problems, including those about substance use. In some cases, this was reported to have led to greater insight. Some people valued the practical work and training support from Astep.

Negative responses focused more on the practicality of INTEGRA contact, and on its relevance for some attenders. Thirteen per cent of responses indicated that travelling on public transport to office-based appointments was difficult for some people; some people did not respond positively to office-based appointments for other reasons:

You never know who is listening in a probation office.

Didn't want to bump into another scaghead in the waiting room.

Moving from room to room each time – not set up for interviews.

Table 6.11: Balance Sheet analysis from T1

Positive		Negative	
Rapport/Trust/Self-esteem	39%	Office/Waiting Area	20%
Hope/Motivation	37%	Relevance	20%
Direct Work or Training Help	36%	Travel	13%
Insight	33%	General	13%
Helped to Talk	31%	Had to do it	7%
Direct Alcohol/Drugs Help	28%	Rapport and Trust	7%
General	21%	Initial Confusion	5%
Information Giving	13%	Probation Generally	4%
Problem Solving	5%		
Impact on Offending Attitudes	5%		

For those alcohol or other drug users who did not consider themselves as 'having a problem', or who felt they were already receiving adequate help, the relevance of their INTEGRA contact was another source of negative response:

Didn't have a problem with alcohol or drugs – one offence when I was drunk a few years ago.

Already getting help with my drugs – didn't see the need to see someone else.

The Balance Sheet feedback was also useful in indicating issues for more detailed enquiry at T2. Below, a selection of clients' views is presented in relation to each of the themes addressed in the second interview. They have been chosen to indicate the breadth of responses.

“Offenders who find a job are much less likely to re-offend”

28 of the 45 respondents (62%) generally agreed with this statement. Some said it gave a structure to the day, and engaged them with the rest of society. Money brought new possibilities, some responsibilities, self-respect and a distraction from the lifestyle of substance use. On the other hand, eight (18%) did not agree. For them the nature of the offending is an important consideration. An expensive drug habit, for instance, or a low paid job, may still require the additional funds of acquisitive crime. Having a job, too, may have little impact on public order/violence offences, or on those offences directly associated with substance use:

I think that's crap. Why? Say you're a junkie or a scaghead jabbing yourself right, that costs a lot of money. Whether you're working or not you're still having to find that extra bit of money to fund that kind of habit, if you know what I mean.

The remaining nine of the 45 (20%) felt the picture was more complex. For them, “it depends” on such things as: the nature of the individual; their drug use; the type of offending; and the nature of the job.

“It doesn't matter what sort of job this is”

Only seven (16%) felt that the nature of the job did not matter. For them, any job helps by bringing structure and distraction, and, if low paid, it could be a ‘beginning’ which, for some, might be supplemented with ‘fiddle work’¹², or by a working partner or relative. Whilst some respondents viewed an interesting job as the ideal, particularly those intent on tackling their drug problems and overcoming boredom, by far the most significant factor in considering work was money:

A boring job is alright if you're getting paid enough. And if you don't get paid enough but it's interesting you can always get some fiddle work if you're short.

¹² My use of the term ‘fiddle work’ denotes work for payment without declaration to Income Tax or Welfare Benefits authorities. I also use the term ‘informal economy’ in this context too, but prefer the former as it tended to be the preferred term of respondents. It implies no judgement on my part of the relative seriousness of the fraud entailed, but allows us to see the activity from within the world view of the respondent.

How much money you get is important - they could just look at it and say why I could be making more doing other things you know - like thieving and things like that - and you lose your freedom with work.

You're going to have to pay them over £200. There's no point any criminal going to work for less than £200 a week. I wouldn't get out of bed for less than £200 a week - even when I just got out of jail. You've got your rent, council tax, TV licence, water rates, phone bill - you might end up really working for £20 a week - what's the point. I mean this minimum wage thing - its not worth wiping your arse on is it? To get them to work you've got to beat whatever they're getting on benefits plus whatever they're getting on the side.

"It's unrealistic to expect offenders to accept a wage which is not more than they receive on benefit." (At this point in the interview I also systematically asked each respondent about their attitudes towards, and activities in, the informal economy.)

Thirty one of the 45 respondents (69%) agreed with this statement. However, some unemployed respondents were unclear about their benefit entitlement out of work, let alone their benefit entitlement if they were in work. The picture is further complicated by the fact that many were able to make additional income from the informal economy ('fiddle work'). Asking for help from formal agencies with this calculation, therefore, and not least from those agencies within the Criminal Justice System, may be felt to be a risky venture. Given such uncertainty, some people found reassurance in the certainty of the benefit system. Housing costs were a source of concern, and the benefits entitlement calculation for some 'absent fathers' was further complicated, too, by the Child Support Agency (CSA), and the additional monies they feared they would lose if they went 'formal':

I occasionally see a job advertised - I think I could do that, I could stand a chance for that. But then there's all the shit that's got to be sorted out beforehand - that's a big part of it you know, your benefits, how much a week would you be better off or worse off - and then there's the money you get from fiddle work. I work in pubs and things but you can get into trouble if you're found out. Words like confidentiality are fine, but it's not as cosy as all that really - at least where Probation is concerned.

For those actively seeking to overcome a drug problem, formal work was preferred to 'the fiddle' as it offered more definite structure:

I don't really care about money - well, so long as I've got enough to get by - it's helping me to keep off the drugs. I could get some fiddle work if I wanted but I'd rather have a 9-5 job - something that's going to occupy my time. I've seen me say I was going to do some work for somebody that day then I can't be bothered and get bang onto the drugs instead - and settle down for the day.

Altogether, 16 'unemployed' respondents (36% of the sample) were active at T2 in the informal economy. Three of the 9 respondents in formal work were still active in the informal economy to supplement their earnings. One of these got their full-time job after 'fiddling' for the company for a while. Of the remaining 20 respondents not active in the formal or informal economies at T2:

- 1 Was in prison;
- 6 Felt unable due to sickness – and got extra benefits as a consequence;
- 4 Had recently moved area, had no contacts, but would do 'fiddle work' if offered;
- 2 Felt unable to work due to child care commitments;
- 7 Lived with working parents or others who helped financially.

"What about the work/training help you received from the Project?"

Agency records indicate that a third of respondents at T2 were not referred to Astep, and some people had difficulty recalling what help they had received, and from whom, whilst on probation. Negative comments seem to centre around dissatisfaction as a result of no direct work or training output, although for some these may not have been wanted in any case. Positive comments reflect Astep's input, which consisted of help in filling out job applications, writing CVs, contacting employment agencies, driving to interviews, and generally engendering confidence.

Some recognition is due to the effects of Astep that cannot be quantified in work-related 'outputs'. For some people this may be experienced as frustrated expectation, for others being exposed to a serious consideration of work in a supportive environment may boost confidence and self-esteem:

Mind you, I suppose thinking about it, Astep did help a bit. I mean you'd go in and you'd think about work, think that you could get a job, that it was possible, that it might help.

The issue of disclosure of criminal records affected the choice made by many people about whether to be seen to be supported by Astep:

Helped with a CV. But I got the job myself. Didn't want to be seen with Probation in tow either. I mean you've got no chance then. Just don't tell anyone about your past.

One respondent had been honest about their criminal record:

I did apply for a job recently and I confessed to having been in trouble - on the form when they ask you know. The personnel woman was very nice about it and thanked me for being honest. But she said straight out that she couldn't give me a job because I'd been in trouble and it was all recent really. She did say mind that in a year's time they might look at me in a better light. But I suppose that's put me off trying really. It was embarrassing you know.

However, the majority of respondents in work had not been open with employers about their criminal history, and approaching employers with Probation/INTEGRA "in tow" was not conducive to such deception. Indeed, 8 of the 11 respondents at T2 listed as an employment output in agency records believe they got their jobs without any Project work-related assistance:

No one has helped me get a job. I got it off my own back. I said to myself right I've had enough of this so I went out and found work. I wouldn't want employers to know about my past anyway. That's how I got my job with the council. I mean I told them about my driving offences and stuff. But I didn't say anything about drugs and burglary and theft and all that. I mean there's no way you'd do that if you want the job.

"How has contact with the Project affected your drug/alcohol use?"

Twenty six of the 45 respondents (58%) recalled their contact with the E Team positively in relation to help with their alcohol and/or other drug use. Interestingly, for some, the benefits of INTEGRA contact seem to have come with hindsight. For example, one person expressed regret that he hadn't felt able to exploit the insights gained from Project attendance more fully at the time:

It switches switches in your head and you know what you're doing then. I used to hear myself saying it - and they would say: "Well how do you feel about doing all that?" - but at the time I didn't

really give a shite - but now it's all over and I think back, I think: Why did I say the total opposite to what I was thinking and doing?

Another emphasised the significance of the relationship he had made with the E Team worker and explained how this differed from his previous contacts with the Probation Service:

*The thing that stands out for me was this life graph he [the Project worker] did with me. You know: 13/14 getting into bother, then 21 it stopped for a while... And he was such a nice lad - not like my Probation Officer. If he'd had done it with me I'd have just knacked it off - I didn't really get on with him. I got on with him to his face - just to get what I wanted, you know what I mean - just to get off Probation early. But with **** I was telling him things I wouldn't dream of telling my Probation Officer.*

On the other hand, six respondents (13%) had a less positive perspective, generally arising from a reluctance to do anything about drug use which they valued, and a resentment at being 'forced' into some sort of treatment:

Just giving you bad things you know. You're going there and you're enjoying your drugs - but they're just telling you bad things. It just don't register you know. When you're enjoying it, you ain't going to stop - not if you don't want to.

For some the impact of other agencies, or even other people, had been important:

Don't remember seeing anyone from INTEGRA. Milestones [Voluntary Drug Agency] were good though - got me out of the house. Wouldn't mind doing that again.

The drugs part [of INTEGRA] was alright. Then I ended up on a therapeutic wing in Holmewood [Prison]. That's what got my head sorted out. All the groupwork and stuff. Learnt a lot of things. Sit down and talk about drugs. Talk about getting work and things.

I just did it myself - well what it was like - when I started working - they knew I was taking drugs - then one of the supervisors gave me a bit of a hand - you know, trusting me and things like that. He was the main help getting me off it - helped me a lot.

Important, too, had been bad experiences with drugs, or simply being 'sick of being sick':

The reason I'm off me drugs now is because of the heroin overdose [2 months ago] - that saved us - I could've died.

She asked me if I wanted to go to NECA [Voluntary Drug Agency] and that. And that was it really. I stopped drinking myself really. I realised that I was getting into trouble when I'd been drinking - you know, fighting and things. .

For others, a significant motivator was their family, and a developing sense of responsibility:

I spoke to him once and that was it. Didn't offer me anything. But I just got to the stage when I thought it was time to grow up and make something of myself. So I knocked me drugs on the head. I mean I've got three children - and I thought: "It's time to move on" - I mean I'm no showroom for them am I? - being a criminal..

"How has contact with the Project affected your offending?"

Eleven people (24%) felt that contact with the Project had some effect on their offending by providing insight:

They helped me see how my drinking was getting me into trouble. I've had enough of all that - not getting back into trouble again. No way. I've had enough.

No bother for me no more. Probation and Integra does help like - they give you a different way of looking at things - so it does help you, but most of it has come from me.

By far the majority of respondents (29, 64%) attributed any changes in their offending attitudes or behaviour to other influences. A fear of further sentencing was one such influence:

It was more my last sentence which sorted me out like. Just got sick of it. I'm not going to prison again. I've had enough. I'm sick of all that. Sick of being in bother.

But the most commonly reported reason was a sense of responsibility for others, and mainly for their families:

I've got four reasons why I wouldn't get into trouble again - my family. I've got far too many things to lose. I've worked too hard to risk losing them and if that meant staying in the house 24 hours a day to keep out of trouble then so be it.

The fact I was going to get locked up if I got caught again - that was the main fact. No - I wouldn't say that was the main fact - I couldn't be bothered with my kid growing up knowing that there was people knocking at my door to buy drugs and stuff like that - that sort of done my head in.

Didn't have much impact on me at all [the Project]. But then I was by myself and didn't worry about getting in a state and getting into trouble. Now I'm planning to get married and that and I've got to change things. She doesn't want me to sit about drinking all the time. Got responsibilities now.

However, five respondents were less 'certain' whether their attitudes to offending had changed in any way whatsoever.

OUTPUTS AND OUTCOMES

The Programme Logic Model for the INTEGRA Project (Chapter 3) lists the output targets at Project inception. These had been decided by the Assistant Chief Probation Officer responsible for the original funding application. I understand from Astep management that the work-related targets were based on figures being obtained by Astep at that time for 'regular' probationers. However, at midpoint in the Project it became clear that targets in relation to NVQs and Key Skills would not be achieved, and were consequently revised down with the permission of the ESF Support Unit. Nevertheless, these remained an enduring problem. Indeed, the Home Office nationally has established a Pathfinder Project to look at the area of accrediting such skills within the performance of Community Supervision. Table 6.12 summarises work-related outputs at inception, as revised down, and what was achieved at Project end.

Table 6.12: Work-related outputs

OUTPUT	INITIAL	REVISED	ACTUAL	% Target Achieved
Action Plans	1000	1000	924	92%
Employment	150	150	121	81%
Further Education	25	25	9	36%
Vocation Training	65	65	35	54%
Basic Skills	50	50	47	94%
NVQs	50	25	0	0
Key Skills Quals.	300	70	0	0

Given that the Project started 3 months behind schedule because of its late approval, and ran for 21 months not 24, it would seem from Table 6.9 that the Project was more or less on target for its

employment and training outputs. However, the employment outputs above, as taken from the Project's monitoring systems, require some comment. Whilst 121 INTEGRA clients may have found a job during their time with the Project, how much credit the Project can claim for this is open to question. The qualitative feedback from clients described earlier in this chapter, suggests that many believed they secured employment without any work-related Project assistance. Indeed, the majority of successful jobseekers in our sample claim they actively sought to disengage from any identification with INTEGRA as this would hinder their chances of deceiving potential employers about the true extent of their offending or drug-taking:

So I just went down to the Job Centre and got myself a job. I didn't tell them about being in trouble though, just kept away from all that. I'd have had no chance if I'd told them. When it says have you got a criminal record and that, you've just got to put "no" - if you want the job I mean.

As for the Project itself, and as evident from the qualitative exit interviews with Project staff explored later in this chapter, there was a constant pressure throughout to achieve its outputs. Such 'output anxiety', I believe, resulted in a temptation for the Project to accelerate 'clients into outputs', and to claim any successful jobseeking as an output, regardless of Project contribution to this. One person in this evaluation's sample is listed in the Project database as just such an output when the 'job' he obtained was actually 'fiddle work'.

As can be seen from Table 6.12, the major shortfall in output targets was in the lack of accredited qualifications. As mentioned above, the original targets were based on Astep's work with 'regular' probationers. No account was made in those targets for the added problems associated with their alcohol and/or other drug use which INTEGRA clients brought. In hindsight, I believe the targets for accredited qualifications were naïve given the amount of consistent time and application required by clients for accreditation. In addition the Project worked on an outreach basis helping people stabilise somewhat chaotic lifestyles. This also meant that contact time was limited with regard to obtaining accredited qualifications. I should say, too, that my contacts with those INTEGRA clients who agreed to participate in the evaluation suggested that many had little enthusiasm for achieving qualifications other than, that is, those vocational schemes which may lead to a job and a wage, for example, forklift driving:

There's no point in them daft training schemes mind – or bits of stupid tickets – not unless it's real and you're gonna get a job and make some money, like forklifts and things like that.

Despite the Project not meeting the mid-point revised targets for accredited qualifications, the ESF Support Unit proved understanding and advised the Project to concentrate instead on the additional non-employment-related outputs identified in my Programme Logic Model. The following summary, drawn from the Project's monitoring systems and my observations, demonstrates how these outputs were exceeded:

- Website designed and implemented to aid dissemination;
- Tackling Drugs Manual developed for practitioners;
- Good Practice Manual for Employers co-produced with INTEGRA projects in Northumbria and West Yorkshire (this manual replaced the CD-ROM output target on the Programme Logic Model);
- Conference organised for 80 Regional employers;
- Conference organised for local practitioners;
- Conference organised for 100 local Magistrates as part of the Project's Exit Strategy to develop policy and practice in relation to the implementation of Drug Treatment and Testing Orders (DTTOs);
- Policy for DTTO implementation initiated;
- Auricular Acupuncture Detoxification Service (ACUDETTOX) established. All E Team practitioners trained and qualified as auricular acupuncturists;
- Five E Team practitioners gained Level 4 Qualification in Drugs Intervention from Manchester University;
- Research commissioned regarding Offenders and Basic Skills;
- Central role in the development of a Durham County strategy for drug-using offenders (under the auspices of the County Drug Action Team);
- Attendance at two National Conferences and one European;
- Twelve exchange trips undertaken to transnational partners, and one transnational conference;
- Regular exchange of materials and experiences with transnational partners;
- Variety of smaller dissemination and training events undertaken, along with the regular production of promotional materials.

But, returning to this evaluation's focus on beneficiary outputs, along with those outcome indicators which may have some bearing on the ultimate Probation Service outcome of 'reducing reoffending':

- the quality of the working alliance achieved with the E team staff and their clients (output);
- the quantity of services used during the intervention period (output);
- whether a client's score on the work behaviours and attitudes scale improved (outcome);
- whether or not a client was in work at the follow-up interview (output);
- whether their self-esteem improved (outcome).

Taking each in turn:

THE WORKING ALLIANCE INVENTORY

In general, scores on the Working Alliance Inventory (WAI) indicate that most clients had positive relationships with their E Team workers. The overall mean score was 215 out of a possible 252 (85%). The standard deviation was 29 (95% CI 208, 222).

Broadly, clients believed that:

Talking with my worker gives me new ways of looking at things;
I am confident in my worker's ability to help me;
We agree on what is important for me to work on;
My worker and I respect each other;

For both cohorts, the most frequent response, was 'very often' or 'always', the only exception being for Cohort 1, who less optimistically rated 'often' for:

I feel the things I do with my worker will help me to carry out the changes I want;

There was no difference between the scores for the clients of different key workers on the WAI. This indicates that E Team practitioners were similarly experienced by their clients, and that, generally, there was a consistent service response to clients, regardless of allocated worker.

However, the mean scores for clients in Cohort 2 (226) was significantly higher than Cohort 1 (206). There were systematic

differences between Cohort 1 and Cohort 2. Cohort 2 was significantly more positive about 17 of the 36 statements that make up the questionnaire. This indicates that, in general, clients in Cohort 2 formed stronger relationships with their key workers than clients in Cohort 1.

The differences found between Cohorts 1 and 2 can largely be explained by the influence that Case Managers had over the characteristics of the first study group when they referred people already on their caseloads to INTEGRA. The *automatic* referral procedure to INTEGRA for those probationers who were unemployed AND had alcohol/drug problems took some time to infuse the CDPS operational culture. As suggested by INTEGRA staff in their exit interviews later, there was the possibility that Case Managers "dumped on" the Project in the first few months of operation by referring their most difficult probationers to gain themselves relief. Whilst being seen by an INTEGRA worker, probationers did not have to 'report' to their Case Managers. More generously, Case Managers in the early days of the INTEGRA Project could not be expected to be fluent with the objectives and methods of the new service. Less than optimistic referrals could be anticipated. Likewise, Case Managers for Cohort 1, unfamiliar with INTEGRA, could not be expected to fully 'prepare' their probationers for INTEGRA attendance.

On the other hand, it should also be acknowledged that INTEGRA staff, themselves, whilst generally being experienced drug workers, were also new to this venture at the outset and may have taken some time to become adept at service delivery.

Given that clients in Cohort 2 had, in general, more positive relationships with their key workers, we might suppose that these clients would also have better outcomes from the Project. I will examine that hypothesis later.

SERVICES USED

Table 6.13 shows the numbers and average length of client contacts with INTEGRA (both E Team and Astep) for each cohort throughout their involvement with the programme, and based on INTEGRA records. It should be noted that these data indicate that some people had 0 contact with the E Team. For Cohort 1 this is due to incomplete data entry in the early weeks of the Project when staff were becoming accustomed to data return and entry systems. This, in part, could account for the lower mean score for E Team

contact in Cohort 1. For Cohort 2, although the minimum number of contacts is 0, the minimum number of contact hours is one. Once again, incomplete data for a number of contacts is the reason.

It must also be noted that not all clients were referred to Astep, and that there was a wide variation in the amount of services used by people in Astep (0-11 hours) and in the E Team (0 to 18 hours).

Table 6.14 shows the amounts of all services used by INTEGRA clients at T1 during the previous three months to evaluation contact, including Probation Services other than INTEGRA. Apart from contacts with their GP, very few INTEGRA clients used any health or social care provision. They were more likely to have contact with criminal justice services. One in five respondents admitted to contact with the police, as frequently as seven times, and several had appeared in court.

The number of police contacts (recorded separately and not reflected in Table 6.14) was highly correlated with mean minutes per week with a Probation Officer ($r .455, p<.001$). This would be consistent with Probation Service practice of tariffing probationers according to a risk of re-offending assessment. Those most at risk are required to 'report' more frequently, and are likely to have more police contacts associated with their criminal profiles and lifestyles.

Perhaps most interesting is the association between the mean minutes spent per week in INTEGRA and the time spent in receipt of health and social care ($r.307, p<.01$). This analysis implies that contact with INTEGRA increased average use of health and social care. This would be consistent with the holistic therapeutic stance of the E Team in encouraging clients to seek appropriate assistance in tackling the range of health and social problems that can be associated with problematic substance use. As such, this finding should be regarded as a positive outcome, even if actual client take-up of health and social supports is low.

Table 6.13: INTEGRA contact

	C1-39 C2-36	Mean	Std. Deviation	Min.	Max.
Number of Contacts	Cohort 1	4.90	4.35	0	25
Number of Contacts	Cohort 2	5.28	4.14	0	18
Total E team Hours	Cohort 1	4.13	3.56	.00	14.50
Total E team Hours	Cohort 2	4.59	3.89	1.00	18.00
Total Astep Hours	Cohort 1	3.29	3.29	.00	11.00
Total Astep Hours	Cohort 2	.78	1.49	.00	6.50

Table 6.14: All services used, mean minutes per week (N = number of contacts recalled in previous three months)

	N	%	Min.	Max.	Mean	Std. Deviation
Probation Officer	65	28	.00	60.00	11.49	13.13
Community Service	7	3	.00	664.62	16.85	86.46
E Team	29	13	.00	60.00	8.50	15.37
Astep	13	6	.00	18.46	1.27	3.56
Other	6	3	.00	13.85	.62	2.38
Psychiatrist	1	0.5	.00	6.92	.09	.79
Community Psychiatric Nurse	1	0.5	.00	6.92	.09	.79
Social worker	3	1	.00	9.23	.17	1.10
Physiotherapist	1	0.5	.00	50.77	.67	5.86
Counsellor	3	1	.00	60.00	1.23	7.60
GP	36	16	.00	15.00	1.25	2.65

It should be noted that the Client Service Sociodemographic Receipt Inventory (CSSRI) used to gather information relating to service use relies on client recall of services received in the previous three months. Some people were more adept at such recall than others.

It should also be noted that client contact with the INTEGRA Project is lower than that with the Probation Service generally, despite normal probation 'reporting' contact being suspended whilst in contact with INTEGRA. This is due to the fact that many people had ceased INTEGRA contact and, where still on an Order, had resumed 'normal reporting' by the time the evaluation interview took place.

The low take-up of health and social supports may, in part, be explained by the 'contactability bias' in this evaluation. Those INTEGRA clients not interviewed may have a higher take-up of health and social supports associated with greater problems. However, the low take-up should also be noted as a possible indicator of 'social exclusion'.

WORK BEHAVIOURS AND ATTITUDES SCALE

As described in Chapter 4, the Work Behaviours and Attitudes Scale (WBAS) comprises 29 questions which are scored on three sub-scales: pride and independence; work anxiety; and work performance. In addition, there is an overall score for the entire WBAS scale.

Examining the subscales, the highest mean scores were for 'work pride'. Thus, clients were more likely to agree that:

*I set high standards of work for myself;
I am able to do difficult tasks;
I can find a job myself.*

Taking both cohorts together at T1 (75 cases), it was found that the work anxiety scale, alone of these four, had unsatisfactory reliability (Cronbach's alpha .58) and this was even worse based on the 45 cases for whom there was data at T2 (alpha .32).

The two INTEGRA cohorts did not differ significantly in the overall WBAS score (total for both groups 94, standard deviation 9). Clients in Cohort 2, in general, had higher scores for 'work pride' (Cohort 1 mean = 3.08 SD 0.35 vs Cohort 2 mean = 3.28 SD 0.41). This would suggest that clients in Cohort 1 were less inclined to recognise the benefits of being in work beyond the financial, had fewer positive work experiences, and were less inclined, perhaps, to jobseek with confidence.

Both scores are higher than the norm published by Harris *et al* (1993) who studied a group of people with long term mental health problems for whom the overall item mean was 2.94 (SD 0.40).

The remaining three scales were found not to be associated with whether or not a person had ever worked.

IN WORK AT FOLLOW-UP

Table 6.15 shows how many people were working in each cohort at the time of the first interview and at follow-up, approximately 12 months later. More people in Cohort 2 had experienced work when they were first interviewed.

Taking both cohorts together, there did not appear to be any increase in the percentage of INTEGRA users currently in work between T1 and T2. However, it appears that the percentage of people with work experience (as opposed to being in work) increased between T1 and T2. This was mainly due to members of Cohort 2 gaining some work experience during the intervention period.

It should be noted that at T2, and as detailed in the qualitative interviews with clients earlier, 16 of 36 'unemployed' respondents were active in the informal economy and undertaking regular 'fiddle work'. Thus, the true percentage of INTEGRA users who were 'working' at T2 may be as much as 56%. The extent of INTEGRA

client involvement in the informal economy is one of the major findings of this study.

Table 6.15: Working status by cohort and interview

Time 1	Cohort 1 (39)	%	Cohort 2 (36)	%	All (75)	%
Never worked	6	15	2	6	8	11
Worked in the past	27	69	25	69	52	69
Currently in work	6	15	9	25	15	20
Time 2	Cohort 1 (23)	%	Cohort 2 (22)	%	All (45)	%
Never worked	2	9	0	0	2	4
Worked in the past	16	69	18	82	34	76
Currently in work	5	22	4	18	9	20

EMPOWERMENT

Self-esteem and empowerment were measured in this study.

Generally, clients were likely to agree with statements like:

I feel I have a number of good qualities;

And disagree with statements like:

I feel powerless most of the time;

It is possible that such responses may have been affected by the context: interviews took place in the respondents' homes where they would, presumably, feel more relaxed and confident.

There were no differences between the two cohorts for the overall empowerment scale, nor for its sub-scales, including self-esteem, power, anger, community activism and optimism. Likewise, there was no increase between T1 and T2.

WORK AND SELF-ESTEEM

It may be assumed that there is, in general, a relationship between being in work and higher self-esteem. This relationship was investigated statistically by Dr. Coolen-Schrijner using correlations and regression analyses. I will report these findings and make my own comments below.

At T1, being out of work was associated with lower empowerment and self-esteem, but these correlations did not attain statistical significance ($p=.10$ and $p=.25$).

At T2, however, being out of work was very significantly associated with lower empowerment scores ($t -3.84$; $p<.001$) and with low self-esteem ($t -2.82$, $p=.007$).

The T2 finding is interesting and is what one might expect: having a job makes you feel better about yourself. However, 16 of the 36 'out of work' respondents at T2 were active in the informal economy, and so could be classed as 'in work'. However, the nature of this informal work is often menial and unstructured. It is also work which carries no rights and no security, and cannot be relied upon for regular activity or income. A key aspect of notions such as self-esteem, empowerment, and even social exclusion, is that of dynamics. It is not simply a question of current status, it is not simply about whether someone has a job or an income right now, but it is also about their prospects for the future, and how they feel about them.

However, the question should also be asked as to whether particularly low self-esteem and empowerment scores for those people out of 'formal AND informal' work may have skewed these results somewhat. As described earlier, the remaining 20 people out of 'formal and informal' work at T2 consisted of: one person in prison, four who had recently moved area and had no contacts, two single parents, six people in receipt of sickness benefits, and seven without independent accommodation. Perhaps there are other aspects of these people's lives, other than lack of employment, which significantly impacts on their self-esteem and sense of empowerment?

Correlations

There were some associations between WBAS scales, empowerment and self-esteem, as one might predict.

Work pride was associated with self-esteem ($p<.001$) and empowerment ($p<.01$).

Work anxiety was also positively associated with self-esteem ($p<.05$) and empowerment ($p<.01$). Work performance was less strongly associated with self-esteem and empowerment ($p<.05$), but the overall WBAS scale was strongly associated with both self-esteem ($p<.001$) and empowerment ($p<.001$).

Does employment affect self-esteem?

Using multiple linear regression analysis, what variables had a positive effect on self-esteem were investigated (a more appropriate analysis may have been undertaken using binary logistic regression – see, for example, the evaluation of Leeds Youth Offending Team undertaken by Kazi *et al* (2002)). As predicted by the correlations described above, work pride and work anxiety were found to be positively associated with self-esteem, controlling for work performance and the WAI scales task, bond and goal.

It was possible to explain 21% of the variation in self-esteem by the variables work pride and work anxiety (F 10.94 on 2 and 72 df, $p < .001$). However, this also means that 79% was accounted for by other factors plus random variance.

It is interesting to note that higher levels of work anxiety appear to be associated with higher self-esteem and not the reverse. Anxiety is often taken as a negative trait, but a moderate amount of work anxiety, it appears, may be related to a concern to perform well at work and this in turn is reflected in self-esteem.

Contrary to expectations, moving between T1 and T2, from not being in work to working, did not appear to be significantly associated with an increase in self-esteem. This may be because so few people changed their working status (5 moved from not working to working, 6 moved from working to not working).

It was not possible to predict statistically which clients obtained work from their scores on either the working alliance scale or the attitudes to work scale. But, once again, the small numbers obtaining work made this very difficult.

There was no statistical relationship between the amount of drugs and alcohol consumed (measured in terms of cost) and either self-esteem or being in work.

STAFF PERSPECTIVES**PARTNERSHIP ISSUES**

During the first few months of this evaluation, it became clear, through conversation and observation, that there were problems with some aspects of the partnership elements of this Project, most notably between E Team and Astep operational components. Respective staff were 'grumbling' confidentially with me about the approach of the other. Whilst their appeared to be sound

communication between partners at a strategic and management level, the Project had abandoned its intention of organising joint team meetings for E Team and Astep staff. In addition, whilst glancing through Astep 'returns' on client progress, I was struck by the number of forms which included the comment: "Still using!". Further investigation revealed that the comment referred to current drug use. This indicated a discrepancy between E Team and Astep operational principles in relation to drug use.

E Team staff, as experienced drug treatment workers, and in line with the Developmental Model of Change (Prochaska and DiClemente, 1984) underpinning their approach, were operating from a harm reduction principle: continued drug use does not, on its own, necessarily constitute a problem for an individual if the difficulties associated with their drug use are now felt to have ameliorated and to be more 'manageable' for the individual. Astep staff, on the other hand, and as it emerged, were operating an abstinence principle and expecting the beneficiaries to be drug-free when they were referred from the E Team. This discrepancy, with its operational consequences, had not emerged during Project planning. One of Astep's concerns expressed to me about the working of the E Team was that they were having to refer some people back to them for further therapeutic input as they were not 'ready' for the Astep phase of the Project. Whilst this information was fed back to both E Team and Astep management for action, it also served to amplify my feelings that partnership issues within this INTEGRA Project warranted some systematic exploration.

However, despite a number of attempts to have this sanctioned by the Project Management Group, it did not come to fruition. Rather, it was agreed that the E Team Project Manager would organise a 'Partnership Day' for all partners. It was also agreed that I would help with some of the content planning for this 'Partnership Day', and include a suitable evaluation component. However, in the event, the 'Partnership Day' did not occur.

Similarly, tentative plans to evaluate aspects of the wider transnational partnership also did not occur. However, this was due to the Portuguese Project finishing earlier than anticipated and not being available for the final transnational meeting in Italy. I took the decision that evaluating the transnational elements would not be fruitful if data could only be obtained from two of the three partners.

Given that partnership issues could not be explored in a more formative way, I decided to introduce them instead into the 'exit' interviews with staff at Project end.

EXIT INTERVIEWS WITH INTEGRA STAFF

I undertook confidential semi-structured 'exit' interviews with Project personnel, including three people who had previously left the Project and all those who were still in post at Project end. In all I carried out 14 interviews, four with Astep staff and ten with members of the E Team, both Project workers and managers. The E Team interviews also included two with staff engaged primarily in administrative duties.

To avoid the need for detailed note-taking, all interviews were tape-recorded.

Given the confidential nature of the interviews, particular care has been taken to anonymise the identity of respondents in presenting these findings. Nevertheless, some of the findings highlight differences with regard to staff roles, especially those differences that emerged between views expressed by Astep and E Team staff, and some that emerged between Project workers and Project managers.

"Offenders who find a job are much less likely to re-offend"

There was broad agreement with this statement. Respondents believed that as a source of purpose, identity, finance and structure, a job can help in reducing the risk of re-offending. E Team personnel specifically highlighted the importance of a job in helping to overcome boredom, in their experiences a particular issue for drug users:

The biggest thing that has come up is boredom. It's why many of them struggle - it's hard to find alternatives on limited resources. They get into a rut - both mentally and physically.

Finding a job also sets other things into motion - structure, stability, another peer group, as well as resources that come from having an income.

Agreement with the statement, though, was generally only given with qualification: clients may want a job, as opposed to training, but they want a job that is going to pay more than they receive on

benefits PLUS whatever other income they can get, say from the informal economy:

There's no doubt that employment does significantly impact on offending behaviour. The difficulty is that when offenders come to you they want a job - they don't want training - but they want a job that is going to cover their costs and more - and that is difficult when they may well be in the informal economy or have other means of getting money. So employment doesn't appear to be a viable option in that sense.

"It doesn't matter what sort of job this is"

There was overall consensus that the nature of the job does indeed matter, and should preferably have some sense of value for the offender, either in terms of cash or interest, as well as offering some sense of security. Some E Team personnel expressed concern that interventions with offenders could be counterproductive if such issues were given little priority.

The job does matter. In the context of Integra, for instance, clients could become resentful if pushed towards jobs they don't want.

It depends on the individual's sense of value and their self esteem. If it's a thankless task it's not really going to do something about that. I suppose it's all tied in to the motivational ideas of "critical self esteem" - we are working with people where this is already low.

While recognising the importance of money and interest, Astep personnel, on the other hand, emphasised the role of staff in helping offenders 'find the value' in some jobs, as one put it:

Even a "dead end job" can have meaning for someone if you can help them view it differently - not "dead end" perhaps - but a "foot in the door". The same goes for temporary jobs.

"It's unrealistic to expect offenders to accept a wage which is not more than they receive on benefit"

There was overall agreement with this statement. As one E Team staff member put it:

Drug users are not fundamentally different to me - and why should I do more work for less money. Heroin is a powerful alternative - it's like 'Ronseal varnish' - it does what it promises.

The responses of Astep staff emphasised the importance of a benefit assessment in helping people comprehend where they stand, both out of work and in relation to a particular job/wage:

... sometimes they do experience the "Benefits Trap" - but this too can be framed positively with some people - "a foot in the door". Sometimes you feel you are bashing your head against a brick wall - and the Probation Officer might agree that's the case, but we'll still have a go.

"How have Probation Officers responded to your work?"

Overall, E Team and Astep staff perceptions of Probation Officer (PO) (Case Manager) responses were positive. Some difficulties arose earlier in the lifespan of the Project, and were, perhaps, experienced more by those staff members not familiar with Probation's enforcement culture. Feelings of being 'dumped on' by some POs, and a sense that their skills were initially undervalued by some as a result of their Probation Service Officer (PSO) status, were common themes in the Project worker responses. Staff already known to, and trusted by, POs, felt some concern for their own workload management when POs made specific requests for their involvement:

Initially, there was some suspicion and resentment. It didn't help that we came online at the same time as a major reorganisation in the Department - so there was already a lot of resentment and suspicion.

POs already knew me. Because of this some POs have asked for me personally when making referrals.

However, the demands made of the Project by POs may have been to the Project's advantage in the early months of its operation, when it was imperative to get up and running, as one E Team management response put it:

There was some inevitable "dumping" by POs - but this was also something of an advantage as we had outputs to meet. We had to go at a pace - and right from the very outset when we only had

about 4 weeks to set things going once we had the OK to our application. We couldn't afford to hang about.

"How have probationers responded to your work?"

Staff generally felt they had been successful in establishing rapport and gaining trust. They also felt that many clients saw them as 'preferable' to seeing their PO, responding to the INTEGRA 'therapeutic' stance rather than the Probation 'enforcement' one.

On the whole really positively. I think they're quite shocked when they realise I'm not a Probation Officer: "Can I have you instead cos I don't want to see my Probation Officer". I suppose this is the difference between my therapeutic engaging role and the PO's enforcement. I make it clear though that what we talk about will be fed back to the PO.

However, staff were less certain about how 'therapeutically successful' their work had been, and this could have a negative impact on their role confidence. In part, they believed this was due to some clients feeling 'compelled' to attend:

The ones engaged have responded OK. But there is such an amount of non take-up with people after the first interview that you're left questioning your own ability. Those who you never get to see don't have such an impact. I think it's because POs tell their clients that they "have to" come, that it's part of their Order. Then they turn up, see us, then realise it's not part of their Order, and you don't see them again.

Of course, as outline in Chapter 3 when considering the Developmental Model of Change (Prochaske and DiClemente, 1984) employed by the E Team, automatic referral to the Project for all clients who were BOTH unemployed and have problems with drug use would be likely to result in significant numbers of clients at pre-ACTION stages. Motivational approaches with problem drug users are demanding in themselves:

The clients have responded to the workers very well I feel. Physically of course they have to turn up. Psychologically though is a different matter. It's hard to engage/motivate people who don't need to be seen, don't want to be seen, or aren't ready to be seen.

However, E Team staff believed that the task of motivating clients, and helping them deal with their drug problems, was made even

more difficult by the little contact time available to them. They attributed this to a 'pressure of numbers', and the Project requirement that 'cases' be closed as soon as an employment output was achieved. It seems there was some inconsistency between E Team staff in addressing the closure requirement:

I sometimes haven't reported the fact that a client has got a job because I feel they need more support - not least given the extra uncertainty built into starting something new - it's a scary thing - but getting a job means automatic closure for us. Some people in the team do the same thing - others close immediately. It all feels like some output-oriented push to get people into jobs. But then again, the numbers of referrals are high - closing quickly is all part of the pragmatics.

The previous quotes come from E Team staff and relate specifically to the 'front end' of the Project, they relate to the alcohol/drugs assessment/help. But similar opinion emerged too from interviews with Astep staff, those working at the 'back end' of the Project. For them, additionally, particular problems have been experienced in engaging clients with the idea of undertaking qualifications:

... compulsion can be met with resentment and I think that this has spilled over and I think sometimes they do resent that they're expected to address their problems whether they like it or not. They have not responded positively to the idea of doing qualifications, including Basic Skills, at all.

Before continuing with this analysis of exit interviews, I feel it would be useful to interject with more detail regarding the issue of 'closure' raised above.

Reasons for Closure

Reasons for closure of cases, as recorded in the Project database, varied. For Cohort 1, the most common reason was non-attendance. For Cohort 2, closure was most likely to occur because a client had entered employment or completed their work plan. This may be further evidence of the greater difficulties presented by Cohort 1 clients. Table 6.16 summarises reasons for closure recorded for both cohorts:

Table 6.16 Reasons for Closure as recorded in Project database

	Cohort 1		Cohort 2	
	Frequency	%	Frequency	%
Non Attendance	14	36	4	11
Completed work plan	7	18	10	28
Employment	9	23	9	25
Training Scheme	3	8	2	6
Order expired	2	5	1	3
Referral on	2	5	3	8
Prison	1	3	1	3
Other*	1	3	6	17
Total	39		36	

* other refers to closures where people have moved to another Probation Authority, died, or where no closure reason is recorded in the Project database.

Continuing the analysis of exit interviews:

“Was the automatic referral process helpful?”

E Team management felt that the referral process, whereby all probationers identified from their LSI-r as being both unemployed and having alcohol/drug problems were automatically referred for assessment, was helpful in achieving Project outputs and providing consistency. E Team management recognised, too, that ‘output anxiety’, the pressure felt in striving to meet the targets agreed with the ESF, was a feature throughout the Project lifespan:

The Project was designed around this - and it provided consistency. It was good for outputs too. Yes there was some output anxiety throughout. At the outset, everyone we talked with in other such projects, like West Yorkshire and Northumberland, all felt the same thing - it seemed to be a part of the Project culture. I know this caused issues for some Project workers - notions of quantity vs quality. Perhaps if we'd emphasised the employment side of things more, this might have been better.

At Project inception, and as outlined in my Programme Logic Model, it was planned for the CDPS assessment team to ‘gatekeep’ for the Project, filtering out ‘unsuitable’ beneficiaries before they ‘reached’ the E Team. In the event, this did not occur. E Team management recognised that this led to a ‘pressure of numbers’, which then led to some Project staff feeling that their therapeutic potential was adversely impacted. However, management also believed that the E Team was the correct place for initial assessments to be

undertaken as the CDPS Assessment Team lacked the expertise in the area of alcohol and drugs:

Generally helpful, although we have felt the pressure of numbers at times. I'm aware that the application for funding had the Dept's Assessment Team built into the equation. However, I always anticipated us doing the initial assessment. The skills don't exist in the assessment team to deal with the extent of, and motivation for dealing with, alcohol and drug problems. We did our own gatekeeping and assessment, and action planning with clients. I know different workers in the team may have different views about the pressure of numbers, and how this might inhibit therapeutic time available.

As indicated in the management responses above, most E team Project workers did indeed disagree with the automatic referral process. They believed that this had resulted in a 'pressure of numbers' which had adversely affected the Project's therapeutic potential, as well as leading, in some cases, to inappropriate referrals and time wasted travelling to offices for appointments which did not take place:

Numbers have always been a problem in doing effective casework. A lot came through where they were now clean or working. I suppose it's good for official numbers and targets. But you possibly also end up closing stuff before really ready - and this makes you feel a bit uncomfortable. Then there's been so much time wasted just 'twiddling your thumbs' in probation offices waiting for people to turn up for appointments.

On the other hand, most E Team workers agreed with the management view that they were the 'best placed' staff to undertake initial assessments, rather than the CDPS Assessment Team:

Perhaps the assessment team could've gate-kept for us and saved us some time - but then again, I think we are the right people to do the initial assessment - we have the skills, and this sets the groundwork for more work when it happens - and provides consistency.

Other E Team workers, though, felt that their having to undertake initial assessments limited the time available for intensive working with clients:

We are a bit of a "catch all" - and do all the meaningful initial assessment ourselves. The assessment team only note if there is a problem with alcohol or drugs. They don't do anything to assess the extent of the problem. All of these assessments we do limits the time available for intensive work.

Similar concerns to those expressed by most E Team workers were also expressed by both Astep management and frontline staff. They felt that the 'pressure of numbers' experienced by the E Team 'front end' of the Project was transmitted through to their 'back end' of the Project. This had resulted in clients being referred to them who were not 'ready' to consider employment or training, and who they then referred back to the E Team for more therapeutic input, an approach which they understood could have a negative impact on clients' self-esteem:

We have experienced people who are not "ready" for our service being referred to us - and this led to high drop out rates initially. Then we started sending people back to the E Team for more work. Initially though we found ourselves gatekeeping too. This wasn't good for clients who were being set up for another negative experience and it wasn't good for their self-esteem.

As with E Team workers, Astep staff, too, reported that the 'pressure of numbers' had negatively impacted on the amount of time available for them to work with clients, and resulted in their doing a lot of form-filling and CV work themselves, rather than taking the time to help the clients to do it for themselves:

When I first started working for DISC [Astep's organisational umbrella] we worked under an empowerment model - working intensively with small numbers. This led to some success. Then our numbers increased, and we had to find ways to cope. Latterly much of the work we could offer was simply to help people with application forms, letters and CVs - we even used to take these home and do them ourselves rather than take the extra time to help people do it for themselves. Some application forms now are very difficult and complicated.

Our initial plan of one hour sessions was down to less than 20 minutes by the end - and then we would take work home to do in our own time.

Astep management believed that the automatic referral process was unrealistic, particularly given the extra problems and work

demands brought by drug users. They also believed that the original output targets for the Project were inappropriate, especially in relation to qualifications, and even claim to recall incredulity when the bid for Project funding was accepted. It is also interesting to note their claim that they had informed the bid author that the targets were unachievable prior to bid submission, but that the bid was submitted regardless:

It's particularly unrealistic when you consider INTEGRA clients who probably carry a multitude of problems from their drug use. To be honest, the original parameters of the Project were unrealistic and have caused problems for everyone involved. The drug element makes our job twenty times harder. But it's the output culture in which we operate - it's all very well to say don't worry about it; at the end of the day, if it doesn't really matter, then why have we got it? I think for that type of programme attaching any number of outputs to qualifications is wholly inappropriate. The initial targets for qualifications were totally unreasonable - even if they were marginally scaled down eventually. I did see the outputs in the original bid and I did say at the time: "There's absolutely no way we can achieve those" But then it was submitted, and the author left the authority. There was sort of an 'in joke' really that this would never be accepted, and then we all got a horrible shock when they rang up and said the bid is alive - we all went: "WHAT!"

"Were referral criteria of drug problems AND unemployment helpful?"

All of the people referred to INTEGRA had to have both a problem with alcohol or drugs AND be unemployed. Generally, members of the E Team and Astep found problems with these joint referral criteria. Whilst they recognised that the issues are intertwined, that finding a job is made more difficult if you have a drug problem, and that there is some consequent theoretical validity, they had concerns for those people who, on finding a job, were no longer eligible for alcohol/drugs help:

Created problems from the outset - inequality of service - some people were barred from receiving the help. If they got a job, or if they had a job, they got no help. I suppose I had to continually remind myself that this was an employment Project, and sometimes this clashed with our service perspective.

I think it is all a bit naive really - the idea that getting a job will somehow take away all of these other problems.

In addition, one E Team worker pointed out that 'getting better may not pay'. Clients in receipt of additional sickness/disability benefit monies as a consequence of their substance use may experience financial disincentives in considering any return to work:

...GPs put people with drug problems onto disability [sick note provided so that people can claim sickness welfare benefits]. They get used to more money and a lifestyle where they don't have to get up on a morning. I once had a client who relapsed just before he was due for a benefits review - he told me he did it on purpose so his money wouldn't be affected.

Astep opinion echoed concerns already expressed in relation to 'job readiness' and pressure of numbers:

Theoretically should've been a useful thing to do. But there were problems with job readiness of some people coming through and the overall pressure of numbers.

I don't know why we just keep on running the usual Astep model. The complexity brought by drug users needs true customised and intensive support - not pressure to accelerate people through the Project to achieve outputs.

Only one response, from E Team management, offered a perspective on why the joint referral criteria may have been helpful. The employment criteria, they believed, helped in caseload management by controlling numbers. They also felt it was 'good' to offer a service with a tangible 'end-point', an output:

Adding employment into the equation clearly helped to keep our referral numbers down to manageable levels. It was good to be able to offer the service with a positive end point in mind - employment or training.

"How have the partnership arrangements worked out in practice?"

Here, I explore first the issues that relate to staff members seconded to the E Team by partner agencies; then those that relate to the functioning of the E Team 'front end' and the Astep 'back end' of the Project.

On the whole, staff secondments appeared to have been experienced positively by the E Team. Secondees themselves felt both a part of the team and yet 'free to think differently', and enjoyed

the additional supervision afforded by having two sets of line managers:

*My ***** status has had no negative impact. I have felt a part of this team as much as anyone else. For me personally, my ***** status has added value. I enjoy not feeling owned by the Probation Service. This has given me an intellectual freedom, as well as the political freedom to say what I think. I have also been supervised to death, which is positive for me. I've had normal Team Supervision from ---, as well as supervision from my ***** manager.*

There was some opinion that secondees may have had difficulty 'settling into' Probation 'enforcement' culture, but that this may also be individual, rather than agency, specific:

I think the seconded workers have had some problems. They didn't understand the Probation role - the notion of enforcement proved difficult for some of them.

*I don't think issues here were so much agency specific as individual specific. For instance, the two ***** workers - the first was very experienced and came to us with preconceived ideas. The next worker was not very experienced, offered more of a blank sheet and therefore fit in better.*

Nevertheless, the E Team management response to the operational aspects of the secondments is clear:

Overall, partnership management has worked really well. And this includes contacts with non-directly involved agencies operating under the umbrella of the Drug Action Team. NECA and NORCARE have been really supportive and responded very professionally. This has added value to the way the Project has worked - for instance, our assessment process was informed by NECA and NORCARE's own work - all of this went into our planning melting pot.

Similarly, both E Team and Astep management saw positives in their INTEGRA partnership at a strategic/managerial level:

As for Astep, there has been good cooperation at the management and admin level - around systems issues and methods.

Our partnership with Probation has added value. I think for Probation it has been useful too - they now have a better

understanding of the funding culture within which we have to operate, and within which they need to pitch their partnership thinking.

However, there was E Team and Astep management consensus, too, that there had been some difficulties with operational aspects of their partnership. These, they felt, were mainly due to a lack of communication and cooperation between frontline staff, and a clash of working cultures:

At the operational level I think there have been communication problems because of inconsistent personnel and some confusion about what we were supposed to be doing.

At Project worker level there really wasn't much collaboration. Initially we set up joint meetings for the workers but these were painful.

We never succeeded in combining our cultures.

It should be noted that E Team and Astep staff operated from different organisational bases. Whilst they worked on an outreach basis, in probation offices throughout County Durham, and would occasionally 'bump into' one another, they had very little physical space or time to 'experience' one another. Initial plans for Joint Team Meetings, I understand, dissolved on the experience of the first two or three which were characterised by a lack of communication and some discomfort for those attending.

Generally, each set of agency workers expressed some dissatisfaction at the other's approach to their work. Astep felt they were sent clients who were not yet 'job-ready', requiring that they sometimes refer these people back to the E team. Continuing drug use by some clients was seen as particularly problematic. It emerged in the early months of this evaluation that Astep operated an abstinence principle, as opposed to the E Team's harm reduction principle, and expected the clients to be drug-free when they were referred from the E Team. This reflected, what they believed was, equivalent employers' expectations, and carried with it a concern for Health and Safety in the workplace, as well as a concern for Astep's professional reputation in making good recommendations to employers:

At the back end you've got to manage the risk very carefully. We have to be as honest with employers as we can be - we don't want

to be referring people to employers and have them disappearing after two days. We need to encourage the clients to be as open as possible - as well as from the Health and Safety aspects. We can't afford to put the client, the employer, or the Project at some sort of risk. I suppose the harm reduction model used by the Probation end is at odds with attitudes of employers very often. This is not something we really had any time to plan for - how we handle that interface with employers when we're talking about drug use.

Some E Team staff, on the other hand, felt frustration at Astep's perceived lack of therapeutic awareness, and perceived lack of innovation:

Astep were useless. They weren't practitioners. Their structures and frameworks were not helpful for communication, and we had different expectations of each worker.

In my view there was a problem with their inertia. This was a brand new task for us, so everything was up for grabs. As far as Astep were concerned, I think they felt they were being funded to deliver a service they were already delivering and a service they hadn't been criticised for - so why should they do anything differently?

I found some of their comments judgmental - I suppose, to be fair to them, the workers had no preparation for the work they were doing with INTEGRA - nothing around harm reduction for instance. And they didn't seek to learn new ways of doing things - "boxed in thinking".

As illustrated in the Programme Logic Model, in addition to the core partners, there existed the possibility of referral from the E Team to other specialist services for therapeutic backup. This afforded them the opportunity to develop a wide range of nominal therapeutic partners.

Table 6.17, shows the referral patterns between the E team and its other partners. Despite the wide range of nominal partners, a relatively narrow range appears in this table, and only two with a therapeutic remit: NECA and Tackling Drugs. Given the 'pressure of numbers' experienced by E Team staff outlined earlier, the question needs to be asked: what would the benefits to the service have been had more 'therapeutic partners' been recruited, or had those onboard already been used more extensively?

Table 6.17: INTEGRA referral destinations as recorded in Project database

	Cohort 1 (n=39)		Cohort 2 (N=36)	
	Frequency	%	Frequency	%
Astep	28	72	21	58
Basic Skills	4	10	5	14
NECA	3	8	4	11
Tackling Drugs	3	8	0	0
Other	1	3	0	0

“How have the trans-national partnership arrangements added value?”

Generally, those E Team staff who visited one or more of the trans-national partners felt the experience to have been of some value. First-hand experience of practice in other cultures, and the impact of differing policy and social structures, prompted staff to question their own perspectives and practice. Self-worth also got a boost by being funded to undertake such trips:

This was a first for us - and everything was learning. We learned a lot about language issues. In a general sense we learned about the practice in other cultures, and about the impact of differing social and political contexts. It was good, too, to get a sense of Euroland values as opposed to the welfare culture that we get from USA - tough love and all that. This was a challenge to our intellectual thinking. Also learned that having 1000 outputs is a bit silly - small is beautiful. So, yes, there were definite positives - it wasn't just a jolly jaunt. You also feel a sense of being valued by being sent to such places.

Astep management opinion, on the other hand, reflected a general lack of involvement in, and awareness of, the development of the trans-national partnership:

We've had little to do with the trans-national stuff. Initially I had great optimism and some of our staff went on a trip to Portugal. I was looking forward to visits, especially with the Portuguese Project, and exchange of information and good practice - I even had plans to establish quicker dialogue via the Internet. Then it all seemed to fizzle out. No more trips were mentioned for us - don't know why.

We had visitors from Portugal at the outset, which was interesting because we seemed to be coming from similar standpoints. There was a plan for us to go to Italy but that never happened. We had

two staff go to Portugal, which they enjoyed and found useful, but then nothing else happened. The whole thing seemed to die a death. Don't know if the Italy trip took place [PB: "It did"] - well, we were never informed.

It would seem from the above that Astep were effectively 'frozen out' of the transnational aspects of the partnership by their 'more powerful' Probation partner. Indeed, I believe this was the case. Whilst I am bound by confidentiality with some information regarding this, I can say that the communication between the two sets of agency workers, Astep and E Team, when they undertook a visit to Portugal was not good.

"What about innovation and blocks to it?"

All E Team responses claimed a commitment to innovation, even to the detriment of the programmes' survival:

We genuinely tried to do something here. The Partnership Management Group was a useful tool - gave us permission to do things differently within a criminal justice framework. We set everything up from scratch - all the principles and all the systems, and all the methodology, and the process was full of integrity. I suppose in the end, with the changes that are happening within Probation - and changes that have happened even since we got started - I don't think the service was sustainable for much longer - our therapeutic approach clashes with the enforcement culture - especially for some POs.

The E Team's sense of blocks to this commitment focused either nationally on the changing nature of the Probation Service as it continually develops a 'control and enforcement imperative', or locally in what they perceived as constraints imposed by a CDPS reluctance to 'take risks'. A commonly stated example of innovation is the ACUDETTOX initiative:

I suppose the acupuncture initiative is a good example. This was knocked on the head initially - I think by HQ really - but someone plugged away at it and it eventually happened. I think HQ thought it might be politically sensitive - offenders getting acupuncture. So I suppose the single most important block to innovation is the system constraint imposed by a Probation Service adhering to National Standards and moving towards a control remit.

Not surprisingly, given the previous sets of responses, some mention was made too of limitations brought by the pressure of numbers:

Time pressures mitigate against changing things. You end up running around trying to do everything for everybody. Rigidity in the application of rules doesn't help innovation either. In principle we are allowed to be innovative - in practice it's difficult to find the time. The biggest problem is pressure of numbers.

Astep, on the other hand, saw little scope for innovation at their back end of the Project. Pressure of numbers and the output culture, again, are central themes, and find particular clout when considering the retrospective 'output-focused' ESF funding process for agencies:

I don't feel there has been any innovation overall. Just a variation on a theme - ie targeting alcohol and drug users. I think the biggest block to innovation has been output pressures - oh, and a lack of initial vision. In the real world, funders only fund certain things - it's hard to be flexible - and they want certain outputs achieved or you will find your funding "adjusted". Our funding works retrospectively based on results - say 8 months or a year behind the actual spending. So we have to bankroll the work initially, or get someone like Probation to bankroll it for us - then we submit our returns and the money comes.

The corollary of retrospective funding, of funding that gets spent before you get the cheque, is that if you don't achieve your target outputs and your funding is consequently 'adjusted', you have to find the shortfall from elsewhere in your organisation.

A very pragmatic concern for Astep workers was the outreach nature of their work, and the fact that their INTEGRA duties formed only a part of their overall workload:

The original model for Astep was one-to-one working - there were just no resources to try to do things differently - say by offering groupwork. Everything was outreach - slotting into whatever office space was available, carting around cardboard boxes of materials. Only about 1/3 of our time was spent on INTEGRA work anyway, so doing things differently for 1/3 wasn't really feasible.

“What are your views on the support and development offered?”

All E Team Project staff reported feeling personally and professionally supported in their roles, with regular and good supervision, and plenty of opportunity, along with the resources, for training and conference attendance. Astep staff, too, reported some positive experiences in relation to supervision, mentoring, and opportunities for training within the company.

“What about team support and development?”

With the exception of two people who felt that the outreach nature of the Project along with changes in Project personnel undermined the potential for team working, E Team staff generally reported feelings of belonging to a team:

I think we have been very lucky - staff meetings, team building days - lots of opportunity for team support. However, we ended up working as individuals really - and I think this was down to the individuals – some people didn't want to do things differently - even despite the pressures of outreach working. Maybe I appreciated the opportunity more - I've worked in the addiction field for 15 years and have seen how little the issue of team working is addressed by other agencies.

Astep staff, similarly, said that this had been encouraged with team meetings and Project staff aiming to be in the office to do their paperwork at the same times each week.

Balance Sheet of Positives and Negatives

I sought to focus staff perceptions of the Project by asking that they compile a Balance Sheet of Positives and Negatives. Their responses were grouped under common areas and are summarised in Table 6.18. INTEGRA staff mentioned more positive aspects of their experience than negative ones. They valued highly the interpersonal aspects of the job, including the supervision, management and teamwork. They appreciated the relatively high level of resources available for the task, together with the administrative back-up. The innovative nature of the programme was mentioned positively by E Team members, as was the transnational aspect and partnership working of their experience. Problems with the local partnership (specifically, operational role conflicts between the E Team and Astep) were also mentioned:

Table 6.18: Summary of issues giving rise to positive and negative comments (N = number of times an issue was mentioned)

POSITIVES	N	NEGATIVES	N
Support/Supervision/Training	9	Context	8
Management style	6	Process	7
Resources	6	Partnership	6
Teamwork	4	Number crunching	6
Admin support	3	Resources	2
Therapeutic	3	Transnational	1
Transnational	3		
Partnership	3		
Innovation	2		
Agency learning	2		

Negative responses particularly focused on the context of the Project within the Probation Service, on the referral process and consequent feelings of 'number crunching', and on operational problems with the E Team/Astep partnership. For example:

Context

Constraints coming from being situated within probation offices and Probation culture;

Attitudes of some Probation Officers;

Process

Flawed assessment procedures leading to too many inappropriate referrals;

Clients not being job ready;

Numbers

Over subscribed – felt like number crunching;

Quantity not quality – high caseload;

Pressure of numbers - leads to frustration and stress;

Partnership

Lack of cohesion with Astep;

Operational communication and partnership;

CASE MANAGERS SURVEY

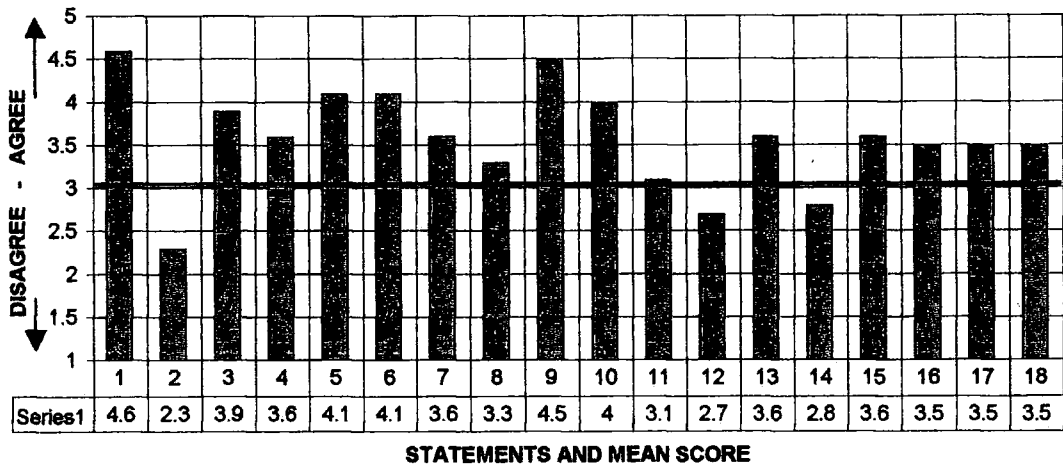
INTEGRA staff themselves suggested that Case Managers (Probation Officers) were major stakeholders in the INTEGRA initiative. Therefore, in October, 1999, 15 months into the programme, a confidential questionnaire was distributed to all Case Managers.

The questionnaire sought to explore Case Manager agreement with 18 statements about offender employment interventions in general, and about the INTEGRA Project in particular (1 strongly disagree; 2 disagree; 3 neither agree nor disagree; 4 agree; 5 strongly agree). These statements covered issues already raised by some officers in conversation with myself, and some issues addressed by other studies (e.g. Samo *et al.*, 1999).

Figure 6.12 shows the average ratings for the 30 Case Managers surveyed. The individual statements are also listed. Among the Case Managers surveyed, who constituted virtually all those Probation Officers who might refer to INTEGRA, there was a high level of support for the aims of the Project, and for the Probation Service's involvement in employment and drug treatment initiatives. This involvement, in their view, should be one of their assessment and then referral on to specialist services. As for CDPS access to specialist services, the views expressed suggested that CDPS had easier access to employment rather than drug treatment services. Case Managers also expressed confidence in the INTEGRA workers' skill base, and were satisfied with the feedback they received from both E Team and Astep phases of the programme. As with the views, generally, of INTEGRA clients and staff expressed in interview, there was agreement that, although a job can help in reducing an offender's re-offending, the nature of the job is important, as is the wage, which should be at least more than what the offender could receive on welfare benefits.

The most notable negative finding is that most people did not think that they understood the working methods of the INTEGRA Project (Statement 14). The majority of respondents also felt that the arrival of INTEGRA had confused the referral process to employment agencies by altering established working practices (Statement 8). I fed these findings back to the programme and it was addressed by them with the E Team Senior Project Worker touring CDPS establishments to re-present the INTEGRA programme at Case Manager team meetings.

Fig. 6.12: Case Managers Survey



STATEMENTS:

1. Unemployed offenders who find a job are less likely to re-offend than those who remain jobless.
2. It doesn't matter what sort of job this is.
3. It's unrealistic to expect offenders to accept a wage which is NOT more than they receive on benefit.
4. Training schemes help people gain the skills they need to secure work.
5. The Probation Service has a key role to play in helping offenders find employment or training.
6. This role should be one of assessment and then referral to a specialist employment agency.
7. CDPS is well-served by specialist employment agencies.
8. The arrival of the INTEGRA Project has confused the referral process to employment agencies.
9. The Probation Service has a key role to play in helping offenders deal with alcohol/drug problems.
10. This role should be one of assessment and then referral to a specialist alcohol/drug agency.
11. CDPS is well-served by specialist alcohol/drug agencies.
12. The arrival of the INTEGRA Project has confused the referral process to alcohol/drug agencies.
13. I am clear about the objectives of the INTEGRA Project.
14. I am clear about the working methods of the INTEGRA Project.
15. The INTEGRA Project workers possess the skills to carry out their work effectively.
16. I am satisfied with the client feedback from the alcohol/drug phase of the INTEGRA Project.
17. I am satisfied with the client feedback from the employment phase (Astep) of the INTEGRA Project.

18. The INTEGRA Project makes a valuable contribution to the CDPS aim of reducing re-offending.

USE AND DISSEMINATION

In Chapter 4, I noted that 'use' and dissemination is a process, and that I was conscious throughout this evaluation of the need to provide suitable feedback, and encouragement for use, on a continual basis. In addition to a number of specific 'events' listed further below, this has involved my:

- Creation of a website – this was placed online in the first few months of my involvement in this evaluation, and evolved alongside. It included the policy and research backdrop to the Project, detailed Project description, the evaluation design and its instruments, findings as they emerged, links to relevant agencies, bibliography, contact details for the Project and the evaluation, and a brief summary of my thesis plans. This website was subsequently adopted by the INTEGRA Project, satisfying one of its target outputs of website provision. The website was also promoted by the European Social Fund as an example of good dissemination practice, and recommended to INTEGRA projects across Europe as a suitable template.
- Regular attendance at, and feedback to, fortnightly E Team staff meetings.
- Regular attendance at, and feedback to, quarterly Evaluation Management Group meetings. Initially, these took place at the end of the full multi-agency Management Group meetings. However, members of the Management Group not on the Evaluation Group felt they would like to be more directly aware of the evaluation and also would appreciate any contributions I may have regarding Project management issues. As such, the Evaluation Management Group was effectively dissolved and I attended the quarterly meetings of the full Project Management Group.
- Informal and regular contact with all aspects of the Project.

More specific events have included:

- Formal interim feedback to, and contributions to accompanying reports for, the multi-agency Project Management Group (with Dr. Justine Schneider) in December 1999, and June 2000. The first of these interim reporting points was built into the original evaluation design

by Dr. Schneider. The second took place at my suggestion shortly after the INTEGRA Project closed.

- Contribution to the formal INTEGRA Support Unit inspection visit on 19/1/2000. I also met with the Support Unit worker the evening before the visit to brief on evaluation issues.
- Presentation to the EMPLOYMENT Thematic Focus Group on 'Empowering the Excluded' in Brussels on 6/12/99. This group brought together representatives from 10 European Members who were focusing on the experience of INTEGRA projects in empowering individuals, groups, and communities. They were specifically interested in exploring 'indicators' of empowerment to develop their evaluation awareness for the next round of EMPLOYMENT funding (2000-2006) which would be called EQUAL and which would have empowerment as a key theme.
- **A Matter of Conviction** – presentation to a Northern Region INTEGRA conference for employers held at York Racecourse on 15/6/99. This contribution was preceded by two visits to discuss employer-related issues with a West Yorkshire INTEGRA project for offenders.
- **INTEGRA and Beyond** – presentation to a conference for CDPS practitioners aimed at local dissemination of Project findings, held at Durham County Cricket Club on 13/3/2000. This conference also sought to sow the seeds for the main exit strategy for the Project and this evaluation's use, the development of strategy and operations for the local implementation of Drug Treatment and Testing Orders.
- **Drug-using offenders: the employment possibilities**, presentation to National Employment Solutions Conference, National Exhibition Centre, Birmingham, June 2000.
- **illuminating the Ethical Space within Evaluation**, presentation to the 18th Annual Conference of the Howard League for Penal Reform, New College, Oxford, Sept. 2000. Sadly, work commitments meant I was unable to accept an invitation to develop themes from this presentation by co-facilitating an ethics workshop at the UK Evaluation Society Annual Conference in December 2000.
- Contributions to the INTEGRA Project final evaluation report – scheduled for publication in February 2001, but eventually published in January 2002.

How the lessons from this evaluation were used by CDPS will be outlined in Chapter 7 where I will also present CDPS response to the evaluation and its report, as gathered from management

interviews undertaken approximately two years after the INTEGRA Project closed.

SUMMARY

At the end of Chapter 3, I summarised the intended INTEGRA Project by constructing a hybrid Programme Logic Model. Figure 6.13, later in this section, is a revised model for INTEGRA at Project end, and is based on the key findings presented in this chapter.

Comparing the two logic models, the first point to emphasise is the absence of the CDPS Assessment Team in gatekeeping for the E Team by undertaking initial assessments. When the Project was conceived, I understand that the cost of the Assessment Team involvement went a considerable way to meeting the CDPS contribution to the 'joint finance' package with ESF. Whatever the reason for its effective non-appearance in the programme process, its absence was felt by staff to have resulted in a pressure of numbers at the E Team front end of the Project. According to staff, this negatively impacted on their therapeutic scope, as did the policy of closing cases as soon as an employment output was achieved. This pressure of numbers could have been alleviated to some degree had the E Team drawn more extensively on therapeutic partners to assist with their casework as originally conceived.

The second point to emphasise is the discrepancy between the E Team's therapeutic commitment to harm reduction, and the Astep expectation that clients would be drug-free when they reached them for job readiness input. This did not emerge during the rushed planning stage. This resulted in tension between the operational aspects of the E Team/Astep partnership which was not tackled. E Team staff perceived Astep as having no therapeutic awareness, whilst Astep became frustrated at the numbers of clients reaching them whom they considered 'not ready' because they were still using drugs. Eventually, I believe, this led to Astep being effectively 'frozen out' of transnational activities by its more powerful Probation Service partner. Given that Astep were the only employment-focused stakeholder at the heart of, what was, an employment-focused programme, this was a significant diminution in the transnational partnership.

Astep staff, too, experienced their own pressure of numbers. In part, this was due to logical transmission from the front end of the Project. However, it was also due to the fact that their INTEGRA duties formed only $\frac{1}{3}$ of Astep staff workload. At the same time,

Astep management were expected to meet targets which they believed were unachievable, particularly in relation to accredited qualifications. According to Astep, and as evidenced in this evaluation, clients, in general, were not interested in training or qualifications, and most only wanted a job if it paid more than they received from welfare benefits PLUS whatever they could earn from 'fiddle work'. However, even if intent on finding a job, most successful job seekers in our sample had actively sought to avoid any association with INTEGRA as this would undermine their capacity to deceive employers about the true nature of their offending or drug use. All of which resulted in significant output anxiety at both ends of the Project.

Whilst under-achieving in some aspects of its work and training-related outputs, the Project over-achieved in relation to its demonstration and network development targets. It also successfully initiated an auricular acupuncture detoxification service (ACUDETOX) which continues today. The ACUDETOX service is the most tangible example of the Project's response to ESF expectation that it innovate. However, innovation is not just about 'doing different things', it's also about 'doing things differently' and overcoming paradoxes and contradictions. In this regard, INTEGRA staff believed that their siting within the Probation Service was a significant block to innovation. This they saw occurring nationally, in the trend towards a stronger 'enforcement' culture, and locally, in what they perceived as a CDPS reluctance to take risks.

As for client outcomes, I believe a quote from one E Team staff member sums it up neatly:

I suppose to summarise: there were those who enjoyed the blag; those who said: "No"; those I saw for a couple of sessions; a lot who didn't turn up; and a few that got some real benefit..

INPUTS	ACTIVITIES	OUTPUTS	OUTCOMES
<p>£2.5 million: Staff Partners Accommodation Equipment Materials</p>	<p>Automatic referral to E Team for all offenders whose LSH notes both unemployment and drink/drug issue. Pressure of numbers and many clients don't turn up resulting in time wasted for E Team staff in travelling to CDPS offices for appointments. Following initial assessment, E Team agree action plan with client if deemed suitable and is willing. E Team work on principle of harm reduction but don't feel have sufficient time for effective therapeutic work. This work could involve other specialist agencies but this is not often the case. Therapeutic work ceases as soon as an employment output is achieved, regardless of stage client is at. Client referred to Astep phase for their help in job-finding and/or training when felt to be 'ready'. Being 'ready' is a source of disagreement as Astep expects clients to be drug-free.</p> <p>IN ADDITION</p> <p>Share knowledge with transnational partners, but 'employment' partner, Astep, not fully involved; Link with other relevant INTEGRA Networks; Disseminate knowledge gained, incl. with employers; Train staff.</p>	<p>action plans 92%; employment 81% (but most clients feel they 'did it themselves' and many seek to avoid INTEGRA identity as that would hinder deception about true nature of offending with employers); further education 36% vocation training 54% Basic Skills Refs 94% NVQs 0 Key Skill Qualis 0</p> <p>IN ADDITION Conferences; ✓ Website; ✓ Handbook; ✓ Staff Qualis; ✓ ACUDETQX. ✓</p>	<p>REDUCED REOFFENDING AMONG BENEFICIARIES (Too early to say, but most clients attribute any changes mainly to impact of their families)</p> <p>ENHANCED SKILL AND KNOWLEDGE IN LOCAL CRIMINAL JUSTICE AND DRUG TREATMENT NETWORKS. (Seems to be the case)</p> <p>ENHANCED UNDERSTANDING AMONG LOCAL EMPLOYERS (Requires assessment)</p>
<p>PREMISE</p>	<p>Offenders who find a job are less likely to reoffend, but the job should pay more than benefits + fiddle work; Training/qualifications could enhance employability but few clients are likely to be interested; Overcoming alcohol/drug difficulties could help if there is sufficient time for effective therapy.</p>	<p>CONTEXT</p>	<p>Probation Service felt to be a block to innovation; Many clients are active in informal economy; Employers are seen as unsympathetic.</p>

Fig. 6.13 INTEGRA Programme Logic Model at Project End

CONCLUSIONS

Firstly, it should be noted that the data which contributed most to the elucidation of the Programme theory emanated from the qualitative approaches and analysis I brought to the evaluation. The quantitative components, however, proved particularly useful in providing evidence of contactability bias.

It is clear that the reliability of outcome data depends on the quality of the sample. Some evidence of contactability bias was found here, which meant that people who were more likely to be using 'hard' drugs, those who had more criminal acquaintances, and those who were less positively disposed towards the Probation Service were under-represented in this evaluation.

The potential impact of a pilot programme like INTEGRA can only be guessed at, since a disproportionate amount of resources is inevitably invested in the setting-up stage. In the case of INTEGRA, setting-up had scarcely ended when winding down began, and so there was never a truly steady-state programme to evaluate.

It is to be hoped that whatever benefits were gained by those involved in this endeavour, service providers and users alike, will have long-term impact, but such outcomes are beyond the scope of this evaluation.

I would like to recommend that any future programmes of this type, programmes which are driven by an employment focus, ensure, at the very least, that they have more 'employment-focused' stakeholders directly involved, for example: representatives from Training and Enterprise Councils, or, perhaps more importantly, representatives from the Chamber of Commerce representing employers.

EPILOGUE

At the end of the INTEGRA experience, CDPS ended its relationship with Astep.

Chapter 7

PRIMARY INTENDED USER RESPONSE

*Now at midnight all the agents
And the superhuman crew
Come out and round up everyone
That knows more than they do.*

Bob Dylan, 'Desolation Row' (song), 1965

INTRODUCTION

This chapter summarises the content of three interviews undertaken during May and June 2002, two years after the INTEGRA Project closed. These interviews were with the three tiers of Probation Service management involved with the INTEGRA Project: the Senior Project Worker, the Team Manager, and the Assistant Chief Probation Officer (ACPO). These interviews sought to gauge their responses to the evaluation, as represented by their reading of the Evaluation Final Report. In my interview with the ACPO, I was also interested in exploring wider issues of evaluation use.

THE PRIMARY INTENDED USER(S)

As described in Chapter 3, whilst having voluntary sector stakeholders, the INTEGRA Project was a Criminal Justice Project, bid for, and operating under the auspices of, County Durham Probation Service (CDPS). It seemed fair, then, to assume from the start of the evaluation, that CDPS was the primary intended user (Patton, 1997) for this evaluation's findings. However, as Patton notes: "People, not organizations, use evaluation information" (1997, p.43).

Patton's "premier lesson" (1997, p. 50) in utilisation-focused evaluation, is that the personal factor directs evaluators to attend to the *specific people* who are the primary intended users. These will generally be the people who care the most about the evaluation.

They will also be people with some organisational power to utilise the findings to influence change.

From the outset of this evaluation, I sought to forge close working links with the occupants of the three Probation Service management tiers involved in the Project. Two of these people, the Senior Project Worker and the Team Manager remained unchanged throughout the Project lifespan. At the ACPO level, though, there were some changes. The original ACPO responsible for the INTEGRA bid, left CDPS before the Project commenced. His successor then took the Project to midpoint before leaving the Service himself. The third ACPO held responsibility until Project end. This Officer has continued to hold CDPS responsibility for drug-using offenders. As such, she has responsibility for Drug Treatment and Testing Orders (DTTOs), the implementation of which, and as mentioned in Chapter 6, constituted the 'exit' strategy for the INTEGRA Project.

DTTOs were introduced by the Crime and Disorder Act 1998 (CDA). They were designed as a response to the growing evidence of 'links' between problem drug use and crime – see Chapter 3. Section 62(1) of the CDA provides that: "...the offender shall submit...to treatment...with a view to the reduction or elimination of the offender's dependency on, or propensity to misuse, drugs". The CDA, then, requires offenders to submit to treatment and provide urine samples for testing.

Whilst the decision to roll out DTTOs nationally was not formally taken until June 2000, the Probation Service centrally had been aware of the likelihood for some time and had encouraged local Probation Services to begin planning their strategies. For CDPS, the experiences of the INTEGRA Project gave them a head start, and both the Senior Project Worker and Team Manager, as well as the ACPO, were heavily involved in drawing up the initial DTTO strategy.

However, since the end of the INTEGRA Project, neither the Senior Project Worker or Team Manager have retained any responsibility for CDPS drug-using offenders. Indeed, the Senior Project Worker is now working for a different Probation Service authority.

INTERVIEW METHODOLOGY

To help focus my interviews with the former INTEGRA Senior Project Worker (SPW) and the former Team Manager (TM), I

decided to adapt a text evaluation methodology commonplace in the field of Technical Writing:

The Plus-Minus Method (de Jong and Schellens, 1997)

This method involves asking respondents, in their reading of a text (in this case the Evaluation Report), and prior to a subsequent interview, to indicate the textual elements to which they react positively or negatively by writing pluses or minuses, respectively, alongside them. The subsequent interview then focuses on the respondents' reasons for their plus or minus marks. The method's openness is its advantage. Readers can place a plus or minus wherever they want, for whatever reason they want, without unduly affecting the reading process (de Jong and Schellens, 1998).

Following telephone contact with SPW and TM, explaining my intention, I wrote to them with a copy of the Evaluation Report and enclosing the following instructions regarding the Plus-Minus method:

Please read the entire report, including the parts that you might usually skip.

While reading, place pluses and minuses in the margins of the report. Write a plus anytime you judge something as POSITIVE – for example: if you find something in the report useful, new information, confirming what you knew already, important, interesting. On the other hand, write a minus anytime you judge something as NEGATIVE.

Decide for yourself which units of the report to mark with a plus or minus. For instance, you could write a plus or minus for a chapter, a paragraph, a sentence, or whatever. It would be convenient, though, if you indicate which part of the report a plus or minus applies to – either by underlining it or by putting a line in the margin.

Any plus or minus is okay, as long as it reflects how you respond to a part of the report. The pluses and minuses, along with your reasons for marking such, will form the basis of our interview.

Interviews, which took place a few weeks later, were tape-recorded with consent to avoid detailed note-taking. At the end of the interview, I retrieved their plus-minus versions of the report to help with my recording and analysis, and left them with a fresh copy.

Semi-Structured Interview with ACPO

My experience of the plus-minus approach with SPW and TM (see later), along with my desire to explore specific areas with ACPO, prompted a different semi-structured approach. In advance of the interview, I wrote explaining that our interview would last approximately one hour, and would focus on a number of questions, thereby allowing her time to prepare for the interview. These questions were:

- Did the INTEGRA Evaluation in any way contribute to subsequent CDPS policy developments?
- From a Senior Officer's point of view, was the one-day ESF Support Unit inspection visit, and subsequent report, more important than the INTEGRA evaluation?
- How did the INTEGRA experience contribute to CDPS implementation of DTTOs?
- How did the INTEGRA experience contribute to any developments in the CDPS response to offender employment issues?
- What were your impressions of the evaluation and the evaluation report?
- Do you expect evaluators to express opinion and make recommendations, or simply to present data?

The interview was tape-recorded with consent.

FINDINGS

Given TM's confusion around the plus-minus method, and SPW's lack of any minus comments other than 'contactability bias' (see later), I have decided to present the compiled findings of their interviews around key themes identified from their responses, indicating differences where they occur. ACPO findings will be presented separately.

At the outset, I should point out a major difficulty expressed by TM in his reading of the Evaluation Report. This was the sense of "dislocation" he felt being asked to undertake this exercise two years after the Project closed:

I suppose the biggest difficulty really was this sense of dislocation from something that, I don't know, was two years ago I guess. [TM]

There is no getting away from the fact that the report's publication in January 2002 rendered it 11 months late. This is an important issue,

and has a bearing, not only on the respondents' ability to recall fluently, but on the evaluation's capacity to inform the CDPS policy-making agenda in the interim. This last point will be explored further in the interview with ACPO.

Contactability Bias

Both TM and SPW felt some discomfort arising from the levels of drug use reported by the evaluation. For SPW, this was not a true reflection of the work undertaken, or of the demands experienced in doing it, by the workers of the E Team, whom she supervised and supported:

You managed to get back to see the people who were less involved with drugs, the less heavy end stuff. What actually the team spent most of their time talking about, or being particularly interested in, weren't those people...they were the cases that they didn't have much success with, the more heavy end...and they don't appear in the evaluation...or not that many do, anyway. [SPW]

For TM, it was even a source of embarrassment:

The bit that I was kind of thrown by was the actual lack of serious drug use thrown up, and particularly as our drugs team now just works with heroin addicts....it was kind of embarrassing that bit. Moving swiftly along. [TM]

The fact that this evaluation's sample of INTEGRA service users was characterised by contactability bias was noted in the report:

...it introduces the likelihood of systematic bias, not just in favour of the more co-operative service users, but also in favour of people using legal substances. (Burlison and Schneider, 2002, p. 25).

Perhaps this bias should have been given much more emphasis? That said, it also raises the important issue of resource 'targeting'. Should the focus of service delivery be on those who are more co-operative, and those that are perhaps more represented by this evaluation? This was a point also noted by SPW, recognising that there is a view that resources should be targeted towards those offenders who are less entrenched in their criminality or drug use:

I suppose it's a personal interest to see if there's anything you can do with people who are more entrenched. But...I mean this fits with Government thinking about who you target; and in terms of

accredited programmes you don't go for the real heavy end, you know. It's not going to work with them, you go for the other end, the lighter end of the scale, and you may have an impact. [SPW]

However, it should be noted that the introduction of Drug Treatment and Testing Orders (DTTOs), with their strong emphasis on treatment coercion and custodial alternatives, was intended to tackle the significant levels of acquisitive crime committed by the relatively small proportion of the drug-using population who are entrenched in their criminality and drug use (Turnbull *et al*, 2000). It will be interesting to see the results of DTTO evaluations in the coming years.

Astep

Both TM and SPW were glad that the report highlighted, what for them, had been the major drawback of the programme, the relationship with, and the operation of, Astep. For TM, he found some relief in the fact that the operational problems between the E Team and Astep were identified in the report, and noted that the Probation Service had chosen to cut links with Astep subsequent to the INTEGRA experience:

And I think it did, in a sense...put its finger on at least one major drawback, and that was the interface between the E Team and Astep...and I think it was a relief in a sense when I read some of the comments about this...

(PB: "The operational problems?")

...yeah. And it was no coincidence that we actually severed our links with Astep. [TM]

TM also believed that Astep's lack of therapeutic awareness, and their lack of innovation, went to the heart of the difficulties:

I think you got into the meat of the thing...the E Team's perceived lack of Astep therapeutic awareness and innovation...it seemed to me that went...if we had a weakness in the Project that went to the core of it. [TM]

SPW, from her therapeutic perspective, particularly noted the comments in the report highlighting Astep's expectation that clients' be drug-free when they were referred from the E Team. She appreciated Astep's position with regard to employers' equivalent drug-free expectations, and the concern for Health and Safety at

work. She also felt that had such information been fed back to the E Team at the beginning of the Project, *rather than the end*, then it could have had a positive impact on employment outcomes:

One of the interesting comments that came out in here was the emphasis on Health and Safety in actually placing offenders who'd had drug or alcohol problems in employment, and the issues that they had about safety and...that was never sort of fed back in the discussions that we had. That would've been incredibly useful if that had been on the agenda right at the very beginning. If that issue had been there at the beginning of the Project instead of emerging at the end of the Project, I think we could've moved a lot further with the employment element. [SPW]

It should be noted that the discrepancy between the E Team's operational principle of harm reduction and the Astep expectation that clients' be drug-free was fed back to the E Team within the first few months of this evaluation commencing – not “*at the end*” as suggested by SPW. The evaluation report was not the first time this issue had come to light. Likewise, I also sought throughout this evaluation to explore the operational difficulties between the E Team and Astep in a systematic way. However, approval for this was never given by the Project Management Group, and their eventual concession of a ‘Partnership Day’, to be organised by TM, and with an evaluative component from myself, never actually took place.

Context

It is clear from the report that most staff felt a major negative for the Project was in its siting within the Probation Service (see Chapter 6). This, they had felt, was a block to innovation. From a two-year on perspective, TMs and SPWs views regarding this had not altered:

...actually being placed within Probation Services was a negative...I didn't find that a surprise... [SPW]

For TM, the situation within the Probation Service generally, and the scope for offering therapeutic approaches, had deteriorated even further:

...even in the two or three years since the Project started the shift in Probation thinking has been much more critical of therapeutic interventions and much more focused on cognitive challenges to people and trying to measure change...and whether that kind of

soft fuzzy wuzzy social work approach...would it survive the harsher life of today?. [TM]

The context for the Project was also characterised by the ESF, with its expectation of outputs and a retrospective funding procedure. This, the evaluation identified, had resulted in some output anxiety and a pressure of numbers.

Output Anxiety and Pressure of Numbers

From his management perspective, TM was open about the 'output anxiety' he experienced throughout the INTEGRA Project, and the need to present "glowing reports" to justify heavy investment, and actual small returns, in working with offenders:

I think the other thing that we grappled with throughout is that any project working with offenders is only going to have very marginal success rates, but that you have to present glowing reports in order to...you know, the kind of output figures we had, in retrospect, were ludicrous...we've (the Probation Service) invested heavily in cognitive-behavioural programmes which are looking at margins of improvement of less than 5%, and that's after a heavy investment of time and energy and hours and looking for those kind of very small margins. Politically we weren't able to...you know if we'd said that we will find 8 people jobs if you give us half a million pounds, that's not an acceptable message...but they're kind of more genuine. [TM]

From her E Team therapeutic perspective, SPW, in a number of her responses, acknowledged the difficulties created by an output focus and pressure of numbers. For example, when talking about the role satisfaction she found in team and staff development, she commented:

I mean one of the briefs was to develop staff and be innovative and that was one of the real pluses of that Project, in spite of the number crunching and everything [original emphasis].... [SPW]

Likewise, when talking about the LSI-r, and how interesting it would be to revisit the service users with another risk assessment tool and make comparisons:

The difficulty was that we had so many numbers because of the original remit. [SPW]

It would be fair to point out here, that during her time with INTEGRA, SPW straddled the sometimes difficult interface between the complaints of team members about the numbers of referrals on the one hand, and the management requirements to generate outputs on the other. At the time, I was impressed by her pragmatic handling of this interface. In this interview, she presented the same pragmatism regarding such matters. For example, when talking about staff responses in the report to the Probation Service context for the Project:

...actually being placed within Probation Services was a negative...I didn't find that a surprise...or the process...part of the process that we had to go through was because of the context...that we were actually stuck within the Probation Service...and the number crunching bit obviously came from the original bid...and the ESF funding...we hadn't got any choice about that...

(PB: "Did you feel restricted by that?")

Yeah...and the partnership bit which I think everybody acknowledged could've been done better...[SPW]

However, regarding the E Team's automatic closure policy for clients who found a job, her reading of the evaluation report seemed to have prompted a rethink:

I mean I think the comment from the person who was honest (about offending) when they went for a job interview (and were rejected as a consequence)...sort of begs the question if you are working in this sort of way that you need to carry on with people longer, you know, so that they're not destroyed by the rejections...[SPW]

Preference for Charts or Quotes?

In responding to the report, TM expressed a clear preference for the qualitative components of the report which had been added to the evaluation core design:

Because of the nature of the person I am, and the experience of the Project, I think what I was more interested in was the more qualitative reflective parts. [TM]

The qualitative components he saw as capturing how the Project had been experienced:

I liked the quotes. They did give a sense...in fact some of the most...bigger lessons actually came through what people actually said...that encapsulated the experience a lot...yeah that's right...so again I did like the qualitative aspects. You got a sense that these people had actually been on the Project and either had or hadn't got something out of it. [TM]

Likewise, TM responded positively to the use of Balance Sheets as a way of focusing and summarising respondents' realities, and providing a snapshot of "what the Project was like":

Clearly having something as stark as positives and negatives (Balance Sheets) does draw out what would seem to be the views of staff...and it felt that did give a good picture of what the Project was like. [TM]

TM's preference for the qualitative components, was reflected in some equivalent impatience with the quantitative aspects of the report. For example, many of the charts he saw as breaking the flow of the narrative, and difficult to grasp:

I just felt it was a little hard to follow...if I'd taken a little bit more time...but as I was reading it...I read it two or three times...and then thought I wasn't entirely sure what it meant. I felt these (Charts) interrupted the flow of the narrative, and I didn't know if there was a different way of doing that...maybe as Appendices, but that may not be helpful either so...there may be no alternative...but they did break up the flow of the narrative. [TM]

Likewise, he had difficulty seeing the relevance in some data and charts, for example those from the Maudsley Addiction Profile (MAP) regarding Health Symptoms and Social Problems:

I lost it a bit on this section (Health and Social Problems) because I wasn't quite clear what it was we were supposed to be drawing out. It felt odd, like it was just thrown in there. Again the charts broke up the narrative. I must admit this bit lost me a bit as well...people being in work or not being in work...I found it difficult to follow...maybe it's just that my brain is deteriorating as I get older. I mean using multiple linear regression doesn't tell me a lot. I would never ever know what that is. [TM]

When I asked TM directly about his preference for qualitative data, he felt that they were the parts of the report that made him think the most, that challenged some of the "trite" assumptions on which the

Project was based, and from where he drew the lessons from this evaluation. He also welcomed reading the comments from clients, believing that we underestimate their ability to "weigh it all up". The complexity brought by the qualitative data, according to TM, demonstrated how the Project took credit for outputs that were actually beyond their control, but that they had to do it "to play the game:

For me, some of these were the most...they were the bits that made me think the most...because what they do is challenge trite assumptions about, well, "getting a job is a good thing, not getting a job is a bad thing". And we underestimate offenders in their ability to weigh all that up. They're not neutral players in this game...they will be making their own informed choices.

I suppose what struck me...you get these like glossies from ECOTEC (ESF Support Unit for INTEGRA) and there'll be people getting certificates with smiley faces and all that, and I'm not knocking any of that but I just think the reality is much more complex, much less politically acceptable. And it just feels like we all have to take part in a kind of pretence.

But to actually make a significant difference is quite hard...you have to understand what it is that they want out of it. So I enjoyed that, that was important stuff for me...and lots about making assumptions about how people do actually change...and clearly we took an awful lot of credit for things that we actually have no control over...because you had to play the game. [TM]

Whilst sharing TM's appreciation of the qualitative components in the report, SPW was not dismissive of the quantitative elements. She found charts useful, and from her therapeutic viewpoint was able to see potential and lessons, for instance, in the MAP Health Symptoms detail which TM had found irrelevant. She believed that such information could be useful for therapists in understanding more fully sometimes why some people have difficulty in achieving their goals. Likewise, she also believed it was important to remember that how someone looks is not necessarily how they feel, and workers need to bear that in mind in undertaking their assessments. If the Project were starting again, she believed that this would be an area she would like to incorporate into staff training events:

It's all very well saying to someone: "How do you feel today?" when they come into a probation office... but if it's more focused I

think that allows them more latitude and I think it also allows workers more...perhaps more understanding of why they may have difficulties in achieving their aims if they were to look at other stuff...

Project workers were always concerned about the physical wellbeing of someone who looked absolutely dire, but if someone came in and they looked relatively alright then a lot of Project workers really wanted to move straight on to other stuff and didn't necessarily recognise that the person...you know, may not have been any where near doing that.

So I thought that was interesting and would be something that...certainly if we were starting again, and looking at assessment processes and perhaps training with regard to awareness, I would want to incorporate into what the Project workers were doing. (SPW)

More generally, SPW viewed the question of 'charts or quotes' as a balance. Numbers and statistics, she believed, were necessary to "justify what you're doing", but the qualitative components pick up a level of detail that the numbers miss:

I think you've actually got to balance it because I think you've got to be able to justify what you're doing, but I think you've also not got to lose the impact that you can potentially have on...you can actually have on people and I don't know that figures always pick up how far people have moved. [SPW]

Overall Response to the Report and the Evaluation

Beyond their shared concern that contactability bias in our sample deserved more emphasis in reflecting the nature of the INTEGRA client group more accurately, both SPW and TM generally appeared to respond positively to the report. SPW had fewer concerns:

There wasn't anything in here that I didn't find useful. No I...there wasn't anything that I found negative either. I thought you had...as somebody who had worked so closely with the Project and been so involved with it there are loads of things that...and there are other bits of work that you have done that I would put together with this...but I think it gives a fairly comprehensive picture...so no, there was nothing in here that I didn't find particularly useful. [SPW]

TM, however, and as described previously, would have preferred fewer charts. He would also have welcomed more conclusions and interpretations:

I think it is entirely appropriate for you to draw lessons and not to just leave us to draw lessons, because we will draw the lessons we want to, and that are right for us. And some of the uncomfortable lessons - we need to learn those as well, and if they're brought to our attention that's better for us. I mean I've drawn out the things that make me feel good about the Project as I've read it, so I would've personally preferred more of that and less of the data.

[TM]

SPW, on the other hand, in considering my recommendation for more employment-focused stakeholders, reflected less certainty regarding the 'evaluator as interpreter' view expressed by TM:

I know it's not your [the evaluator's] place necessarily to draw conclusions about how things might have been done differently...[SPW]

Nevertheless, both SPW and TM agreed with the recommendation that such employment-focused programmes as the INTEGRA Project involve more employment-focused stakeholders:

I thought that was an excellent point there...that we should've had someone from a local TEC or Chamber of Commerce as a stakeholder. That would've, in retrospect, helped our thinking on some of the issues in relation to the realities of the labour market.

[TM]

TM would also have liked to see more description in the report of the formative and utilisation-focused character of the evaluation as it evolved:

I thought this utilisation-focused evaluation (Amendments to Original Design Paragraph of the Report) didn't do justice to what I thought was the extremely valuable aspect and the contributions you made. You know you attended team reviews and team meetings and fed information and opinions back to us, and that ought to be in somewhere...it's not really in this...but it really ought to be somewhere... you certainly made me think... your evaluation evolved and it felt like that was important work that needs reflecting on somewhere. [TM]

As for their abiding memory of the INTEGRA Project and its legacy:

I suppose what's come out from the research (from the evaluation), and I suppose what came back to me in feedback... it's worth investing in staff, it's worth investing in staff development, it's worth involving staff in developing the way they work and in developing the Project. [SPW]

I thought that in terms of a two-year on perspective the impact of the Project was to offer a model that's now currently being used in the delivery of the DTTOs. [TM]

INTERVIEW WITH ACPO

Did the INTEGRA evaluation in any way contribute to subsequent CDPS policy developments?

In ACPO's opinion:

It did and it didn't.

More specifically, as the INTEGRA Project's exit strategy became the implementation of DTTOs, then the evaluation had been helpful in formulating the new service:

...so the evaluation of what had worked with substance-misusing offenders, and what hadn't, contributed to our thinking in the way that we set up the delivery team for the new orders, the DTTOs. So that was really quite helpful....obviously DTTOs are different to INTEGRA but at least it was in our thinking about some of the things that had worked and some of the things that hadn't..

However, the evaluation report, as mentioned earlier, was 11 months late in its publication. Did ACPO feel that it arrived too late to be maximally useful? The answer was: "Yes":

If the evaluation had come out in, say, the July of 2000 it would've been absolutely perfect timing. That was at the point that I had got responsibility for developing Drug Treatment and Testing Orders. But that was unrealistic because INTEGRA only finished in March 2000...but in terms of policy development that would've been the perfect timing.

However, she also noted that there had been interim reporting, and that I had tried to work closely with the Project in a formative way.

The Project, she thought, had developed useful knowledge along the way which was used in formulating the policy for DTTOs:

...we knew we had on the ground feelings about things which had been good, about things which had not worked. We knew we had some experience as a Probation Service of actually delivering treatment, in the widest sense, to offenders who were misusing drugs. We knew some things that had worked so the issues around lifestyles, life skills, constructive use of time, was knowledge that we already had because of INTEGRA – so we just used that knowledge in formulating the policy for DTTOs.

Interestingly, ACPO felt that the evaluation report, post-implementation of DTTOs, was still useful, particularly in convincing more “clinical” DTTO partners that the CDPS approach to ‘treatment’ has some validity. As such, she found it helpful that the report gives credence to such issues as offenders’ self-esteem, skill level, and attitudes to work:

You know we had to fight quite strongly with our partner organisations whose view of treatment was clinical. Well, treatment is different for every individual but it’s actually appropriate to think of constructive use of time as a treatment. And we had experience of INTEGRA to be able to waft in front of them (Picks up evaluation report) and say that this is what we suspect will come out as part of the evaluation. And there are some things in here, because it talks about attitudes to offending levels, skills, self-esteem, attitudes to work and all those sorts of things; they’re things that have come out and we can wander along and waft in front of our partners and say: “Here you go”.

From a Senior Officer’s point of view, was the one-day ESF Support Unit inspection visit (by ECOTEC), and subsequent report, more important than the INTEGRA evaluation?

ACPO recognised that this had been a significant day for the Project and herself. It was the day that they received approval from ESF about the way they had spent £2.5 million:

It was the tick from ECOTEC that says you might have spent £2.5 million but, you know, we’re happy with the way that you’ve spent 2.5 million because we can see some value. And that for me was a big sigh of relief.

You know the spiel was all about that it's innovative, it's supposed to be expensive blah de blah de blah, but, as an Assistant Chief, you're sitting there thinking: "Yes I know, but it seems vastly expensive compared to the things that we normally do". But then ECOTEC came in and talked about the fact that we were achieving results that other areas hadn't seen – other areas hadn't managed to achieve the level of results that we were getting. So yeah, it was almost like a validation that we were out there and we were actually doing the right thing. So yeah, that was a big day.

Not getting "the tick" from ECOTEC could have had serious financial implications for CDPS. The ESF retrospective funding procedures applied as much to CDPS as they had done to a voluntary sector agency like Astep. ACPO believed that the original bid targets, in hindsight were not achievable, and this had been a risk for CDPS in possibly being asked to repay monies, monies already spent:

...the targets, with hindsight, were never going to be achievable. So lesson learned, do a piece of evaluation before you sign yourself up to targets – make sure that you have fully understood and worked out whether or not you can in fact achieve them before you sign the document which says you'll have a go at it... if for example we hadn't changed our targets and we didn't inform them that we weren't going to meet our targets, two years down the line they could come along and say: "Right, well we want a million pounds back, thanks very much". And it's a million pounds that you've already spent...so it's going to have to come from somewhere else.

And this funding procedure applies for five years, meaning that the "spectre of audit" remained:

...so the spectre of audit is still something that the Probation Service has to live with because ultimately they (ESF) can come and inspect...the Project's finished but the financial records are still there and they could still ultimately come and say: "Well, actually, we want some cash back".

Given the possibility of such financial penalties, had ACPO any reservations about undertaking such ESF projects in the future? She did, but also felt that a lot had been learned by undertaking the INTEGRA Project, and a lot of the necessary auditing systems were now in place.

We'd have to think about it. I mean now the systems are in place, we have an understanding of it, so I suppose the lessons and the

knowledge is already there, we would just build on it, so it would be easier the next time around.

Nevertheless, she admitted she would need to think very carefully, and would certainly ensure that forecasted targets were achievable, as well as having a clearer idea for an exit strategy much earlier in the process:

So lesson learned, do a piece of evaluation before you sign yourself up to targets – make sure that you have fully understood and worked out whether or not you can in fact achieve them before you sign the document which says you'll have a go at it... we'd have a much clearer identification of an exit strategy...and how to mainstream it, very very much earlier on in the Project...than what we had.

How did the INTEGRA experience contribute to any developments in the CDPS response to offender employment issues?

As should be evident by this stage in the thesis, CDPS ended their relationship with Astep following the experience of their INTEGRA partnership. ACPO echoed the concerns expressed by E Team staff regarding Astep not achieving targets, and their perceived lack of therapeutic awareness and innovation:

...ermmm Astep from our perception was the piece that didn't ever quite achieve what it should have done...it never ever got anywhere near its targets, it didn't...didn't appear to be engaging with the offenders in a way that we would've wanted them to...and that's why that was one of the partnerships that we let go...

When I put it to ACPO that Astep had been pessimistic from the outset regarding the CDPS-determined targets in the INTEGRA bid, and that they had been unable to devote 100% staff time to INTEGRA duties, there was little acknowledgement:

Astep could quite rightly say that we weren't perhaps very clear in our instructions to them...ermmm...I thought we were, but there you go.

Following Astep's departure, and in the light of the INTEGRA experience, CDPS reviewed its expectations regarding offender employment schemes and negotiated a new partnership. This is a partnership which the ACPO said was working well:

...Tony Walsh Associates have come in and it's so easy apparently...we've piloted the programme in Darlington, and, Community Service as it was, are saying that it's great you know...they just refer refer refer...offenders come back and they talk about how they've been engaged and what they're doing next week.

Although I would not wish to be an advocate for the departed Astep, it is still worth noting at this point that the work of the offender employment scheme has now returned to engaging 'standard' probationers, not the concentrated stream of alcohol and other drug users which populated Astep's INTEGRA workload. One can only speculate what the outcome for their CDPS partnership would have been had I been successful in exploring the relationship between Astep and the E Team during the course of this evaluation.

Impressions of the evaluation and the report?

ACPO's immediate response to this question related to how many clients got jobs, the outputs - that was the significant piece of information she held on to as it was something tangible in justifying the investment:

Because that's something tangible that most people can relate to. So when you talk about 2 to 3 million pounds people say: "Phew that's a lot of money", but then you talk about that that got X amount of people either referred into education or actually into fulltime jobs they can see the 360 degree side...that you've taken somebody who was A/ an offender, and therefore had, you know, an obstacle to get over in terms of employment anyway. Then you add alcohol and drugs to it with accommodation and all the chaos that that sort of lifestyle brings, and you can then say to people and we then got these people into either jobs or education.

It seemed the complexity of the situation, as painted by the evaluation, and as recognised by TM earlier, had either not reached the eyes of ACPO or was not relevant to her 'bottom line'. Beyond the outputs, though, ACPO felt that the evaluation had been helpful, and had provided an anchor point for thinking and reflection:

I found the evaluation helpful because it sort of held us all up and made us stop and think about: "hang on a minute...are we actually going down the right route?" ... you know...or: "should we give this a bit more time".

And we did get reports part way through the process as well, which helped...certainly me...it helped me to pick up because I had this, as you say, I had this thing which was just landed on me...about drugs and offenders and employment per se and suddenly had a responsibility to deliver this Project.

Interestingly, given that she inherited responsibility for INTEGRA at midpoint, ACPO had found the evaluation a source of “some comfort” and support:

So the fact that it had been externally evaluated gave me some comfort. Because even though it might be too late to do anything about it, it would at least come along and say there is somebody else looking at you so there is some support there.

ACPO also believed that other stakeholders had responded positively to the evaluation, not seeing it like an audit, and had been “relatively honest” with me, as well as engaged in the process. She also believed that the evaluation had asked pertinent questions:

I found that most people, because they knew the evaluation was going to happen right from the start, were engaged with it. So my perception was that people were relatively honest with you.

Sometimes evaluation can feel like an audit, and people hear the word audit and “aggghhhh”, you know they’re like cats in the headlights aren’t they? But if you call it evaluation, people are much more comfortable with it.

It could’ve been construed as criticism because you were asking pertinent questions about the way things had been...and effectively coming back to a judgement that could’ve made people uncomfortable...but I didn’t feel that people did...I felt that people were OK with the process.

As I say, for me, it was actually a help because I knew halfway along I had this report which said these things are actually OK. That was good for me.

As for the evaluation report itself, ACPO felt it to be a good report overall. She thought it was well structured and clear, and there was nothing about the report that struck her as being negative. She appreciated the amount of information in it, and was particularly pleased with the LSI-r analysis undertaken as this would provide her with a baseline against which to compare future analyses with a

new risk assessment tool, OASys, which has been adopted by CDPS. The report, then, she felt would have future use:

*I've actually given it to my DTTO manager to go and read and you've just reminded me that she's still got it...
ermmm...yeah...overall a good report...helpful to me...certainly the criminal histories, the tabular stuff was helpful because it's a reminder of what the INTEGRA offender looked like.*

But in terms of future use, for me it's a helpful benchmark for which I can then put OASys against...so, for instance, the LSI-r said this for those offenders and OASys now says this for these offenders...are there any differences, and if there are what are they and why have they come about? So for me that's going to be really quite helpful... there are things in here that I'll be taking – and obviously there's a lot of work in here and you have to do something with it...you can't just have a piece of paper sat there doing nothing. And as I say, we used some of the initial research for DTTOs, and I'm happy with that.

Unlike TM, and with a background in finance, ACPO had a preference for charts and numbers. Whilst she appreciated qualitative data, it was the numbers that would 'sell her an argument'. If she were unsure, or the numbers were a "bit dodgy", then the qualitative might sway the argument. But without any numbers...:

...it would go straight in there (gestures to bin).). It's about effort you know...somebody needs to sell me an argument...you need to tell me...you need to have sold me something. And just somebody over here saying, "it would be nice wouldn't it" ...that is not going to swing it for me...as a policy maker I'm not going to be able to do that...

So if I'm going to receive a report that is going to inform a business, and I'm going to make a business decision on it, then I need something hard, I need something evidential, and the qualitative stuff is great, and as I say, if the numbers were a bit dodgy the qualitative stuff might push it, say, to explore it a bit more or lets have another look. But if it was mainly qualitative stuff, I wouldn't engage with it at all because I would think that somebody hadn't made an effort to sell me something..

ACPO also expected evaluators to make recommendations, seeing that as the role of an evaluator:

If I've got somebody engaging with me and my organisation, I want them to come up with an opinion. And something like this (gestures with report) what would be the point of you just doing the data for me and then allowing me to make my own interpretation of it? Because, with the best will in the world, you will have much more knowledge of that data, I'm relying on you in fact to come up with an opinion and that in my view is the role of an evaluator.

SUMMARY AND CONCLUSIONS

Before discussing the findings in this chapter, I should point out that the three respondents may have felt pressure to give favourable responses during the interviews because of my presence. An independent interviewer may have generated a different response. Likewise, two years is a long time in the lives of busy professionals, and as already noted in his responses, TM had some difficulty engaging with aspects of my interview given the sense of time dislocation he felt. However, this is not just a matter of effective recall across a busy two years....:

...the other problem you have about remembering something that happened two years ago - your memory filters. You remember all the good bits - the midnight strolls through St. Marks Square, and some pretty exciting team meetings and stuff. And that does filter out some of the harsher experiences. [TM]

Nevertheless, what seems clear from this investigation of 'primary intended user' response to the evaluation final report, is that the evaluation did prove useful. It would seem, though, that the use was derived more from my adoption of a utilisation-focused approach throughout, rather than the final report itself.

At the Project level, and as TM pointed out in his responses, my merging of a formative approach with the, generally, summative core design resulted in my, and the evaluation, being seen as a "part of the Project". The evaluation evolved "along with" the Project:

...for me, the evaluation was more of an organic thing which evolved and...you were essentially a part of the Project which for me was really nice. You were someone we could bounce ideas off and think things through so that it felt like it ran alongside...no, not alongside, it ran along with the Project. [TM]

The final report was 11 months late, and, as the ACPO considered, it had passed its point of maximal local instrumental use, at least in

relation to the Project's exit strategy. However, I believe that my formative approach proved to be particularly important. It should be remembered that the core design was experimental and, generally, summative. My manager expected a more distanced stance, focused on data-gathering.

What is evident from this investigation is that the primary use for the evaluation was laying the ground for the implementation of Drug Treatment and Testing Orders. This, though, was more a matter of serendipity than good planning. At the outset, the Project had little idea of an exit strategy, beyond dissemination and knowledge generation. DTTOs proved a handy door for the Project to walk through at the close, and gave CDPS a head start in planning a service to meet the demands of DTTOs in its work with drug-using offenders.

As for the evaluation's local use in relation to offender employment issues, it is difficult to be diplomatic. I am sure the experiences of INTEGRA and its evaluation have 'opened' CDPS' eyes regarding the naivety of their original premise and their associated employment/training targets. Nevertheless, I believe that Astep unfairly bore the consequences of such naivety and its resultant output anxiety and poor relationship with the E Team. Had this relationship with the E Team, and the discrepancy between their operating principles, been subject to a systematic investigation, I feel they may not have been such a casualty of the INTEGRA experience. But all of that is speculation.

What does not seem to be speculation is that all three respondents, at their differing levels of INTEGRA Project management, gained some added benefit from the evaluation. SPW talked of 'therapeutic insights' gained; TM spoke openly about his knowledge gain in relation to the complexity of offender employment issues; ACPO gained "some comfort" from the evaluation when she inherited responsibility for INTEGRA, she found something in the evaluation report itself which had helped in her discussions with more 'clinical' DTTO partners, and she now had data which would prove useful with regard to future CDPS planning.

It was interesting to explore the three responses in relation to their preferences for the kinds of data presented in the evaluation report ('charts or quotes'). TM clearly favoured the qualitative aspects of the report and found the charts "got in the way" of the narrative; SPW appreciated both and felt that the report got the "balance" right; whilst also feeling the report achieved a correct balance,

ACPO, at her senior officer level, and with a background in finance, had a clear preference, as well as a perceived role requirement, for charts and numbers. This emphasises the need to understand the primary user(s) preferences when writing reports. It is not safe to assume, for instance, that all project managers will require detailed charts and numbers, as evidenced by TM's responses. Whilst such differences may reflect differing ontological or epistemological preferences between the respondents, it is equally not safe to assume that either.

In Chapter 2, using Table 2.1, I summarised the basic components of the two paradigm poles, positivism and constructivism. It could be assumed from this, that those stakeholders with a preference for quantitative data, such as ACPO, hold a positivist stance in relation to how they understand the world. And likewise, those stakeholders with a preference for qualitative data, such as TM, edge towards the constructivist end of the paradigm spectrum. This, though, is not necessarily the case. As Hammersley (1996) points out, qualitative inquiry is often based on realism; the view that research is somehow concerned with 'lifting the veil' that covers the 'one reality', thus revealing what is going on.

In a similar vein, it is also interesting to note that, whilst the report was seemingly perceived as well-written, and the evaluation seemingly perceived as well-undertaken, there was a view shared by TM and ACPO that they would have preferred more evaluator interpretation and opinion in the report. SPW, however, was less certain that offering recommendations was part of the evaluator's reporting role. From this, it could be assumed that TM and ACPO have a preference for, what Abma (1997) calls, 'readerly' reports.

'Readerly' reports are characterised as being: "highly structured, controlled, and predictable. Ambiguity is reduced... Giving precise and clear instructions how to read the text" (Abma, 1997, p. 115). 'Readerly' reports are imbued with the authority of the author, and recommendations and judgements are often to be expected.

In contrast, 'writerly' reports are not imbued with such authority, and resonate with the belief that: "evaluation is a collaborative undertaking that confronts us with the social construction" (Abma, 1997, p. 116) of meanings. In this way, audiences interact with the report to re-construct meanings, and recommendations are negotiated not given. Are we to assume, then, that this was the expectation of SPW, but not of TM and ACPO?

Within the evaluation community, as evidenced in Chapter 2, there is a divergence of opinion regarding the legitimacy of evaluators making recommendations or rendering judgements. Some, such as Stake (1982), argue that the evaluator's job can be limited to supplying the data, leaving the stakeholders to render whatever judgements they wish. Others, and most notably Scriven (1986), advocate that it is the evaluator's responsibility to draw conclusions and render independent judgement.

However, as Patton (1997 p. 316) points out: "Utilization-focused evaluation treats these opposing views as options to be negotiated with primary users".

Whilst I share Stake's constructivist hesitation, I also have a strong utilisation-focused commitment to stakeholders finding whatever meanings they can, in whatever way they feel most comfortable, to generate whatever local use is possible. From my constructivist leanings and my utilisation-focused perspective, then, I believe that evaluators must know their audiences and their expectations, they must not assume anything, and they must negotiate the construction of their reports with primary users.

I have to say that such negotiation was not a feature in the construction of the INTEGRA evaluation final report. Rather, the report was 'finalised' by my manager, who has a mix of a 'readerly' style and a 'writerly' perspective that the evaluator's job is to present stakeholders with their data. However, given the closeness of my working relationship, I believed I understood the audience sufficiently to know that any negotiation between myself and my manager to include recommendations would be well-received, and indeed expected, by the primary users:

I'm relying on you in fact to come up with an opinion, and that in my view is the role of an evaluator. It's to evaluate the Project. Did it do what it set out to do? If it did, then say so. If it didn't, then say so. I, as a receiver of that opinion, can then come and have a debate with you about how you've got to that judgement. But just to give me the raw data, I wouldn't have been comfortable with that..

[ACPO]

At the end of the day, the primary users would still construct their own meanings from the report, with due regard to their philosophies, their policies and plans, and their political imperatives, regardless of the evaluators' recommendations.

Chapter 8

DISCUSSION

When human behaviour is the data, a tolerance for ambiguity, multiplicity, contradiction, and instability is essential. When we at last sit down at a clean desk in a quiet study and begin to assemble the vivid images and cryptic notes, searching for a coherency, we must constantly remind ourselves that life is “unstable, complex, and disorderly” everywhere.

Margery Wolf, *A thrice-told tale: Feminism, postmodernism, and ethnographic responsibility*, 1992

INTRODUCTION

The aim of this chapter is to reflect on the entire process of the INTEGRA evaluation, bearing in mind the content of previous chapters. In this way, I will demonstrate how politics, theory, design, philosophy, methodology, ethics, and representational forms, all combine to create the material from which we construct meanings. I will also consider, in hindsight, the extent to which this evaluation provided useful information about the INTEGRA Project.

As mentioned in the Introductory Chapter, and as threaded throughout the text, another objective of this thesis was to discuss a range of methodological issues which relate to evaluation as a general process. It was hoped that integration of such general methodological argument with specific INTEGRA case study reporting would help elucidate the impact of my philosophical and theoretical approach to evaluation as well as identify lessons which can be cautiously drawn for the art of evaluation itself. I will finally conclude with a brief consideration of such lessons.

WHAT KIND OF EVALUATION?

At its initial core, the core I inherited on my appointment to the evaluation, the INTEGRA evaluation design can be described as: quantitative and experimental, longitudinal and summative:

QUANTITATIVE AND EXPERIMENTAL

Influenced by the philosophical postulates of positivism, the epitome of 'good science' is regarded as the inferential statistical analysis of quantitative data, and preferably data that arise from experimental designs. The aim is to 'explain' by demonstrating causal relationships between certain outcomes and the services intended to produce those outcomes. Within the evaluation field, Campbell and Stanley (1963), as outlined in Chapter 2, are regarded as providing the founding text for such 'scientific evaluation'.

As described in Chapter 4, whilst the methodology did not allow the collection of pretest data, with the respondents being interviewed after their INTEGRA participation, the core of the evaluation itself can be very generally characterised as a pretest-posttest experimental design (Rubin and Babbie, 2001). However, as outlined in Appendix I, the term 'quasi-experiment' is sometimes used to describe such impact/outcome designs (Campbell and Stanley, 1963; Cook and Campbell, 1979) where there are no randomly assigned control groups. In the INTEGRA evaluation, an abortive attempt was made to construct a comparison group (see Chapters 4 and 6). A random sample of 'regular probationers' was drawn from the LSI-r database of Probation Service clients in County Durham for the equivalent time periods of our cohorts (see next subsection) and an attempt made to undertake the evaluation interviews with these. In the event, though, three weeks of solid door-knocking, and increased doorstep antagonism towards myself, resulted in only six completed interviews.

However, despite this lack of success in directly constructing a comparison group, we were able to statistically equate users referred to the INTEGRA Project with those who were not referred (the comparison group), on measurable characteristics associated with some of our outcomes, by drawing 'controls' from the LSI-r database.

LONGITUDINAL AND SUMMATIVE

Longitudinal studies involve the comparison of groups at different time periods. The most common approach in attempting to measure the impact of a programme on its participants is to gather pre-programme data, then re-interview the participants post-programme and make comparisons. In the INTEGRA evaluation, however, there was no pre-programme data gathering. Instead, two cohorts of service users were included in the sample: all those people who had been referred to the programme between June and December,

1998 (Cohort 1 Time 1), and the people referred between June and December, 1999 (Cohort 2 Time 1). Some respondents at Time 1 had completed their INTEGRA attendance, whilst others were still being seen. Respondents who could be found, and who were willing, were re-interviewed 12 months later (Time 2).

Scriven (1980) is responsible for introducing the 'summative-formative' distinction in evaluation. Summative evaluations are deductive, have output-focused goals, emphasise evaluator distance from stakeholders, and, generally, produce findings at the very end of the study. These were the characteristics that my evaluation manager had in mind when she designed the evaluation. However, whilst the design had a summative end point, with a Final Report scheduled for publication in February 2001, an interim reporting point was also built into the design in December 1999. At my suggestion, a further interim reporting point was arranged shortly after the Project closed in June 2000.

UTILISATION-FOCUSED

Underpinning the experimental, longitudinal, and summative core, was a commitment to Patton's (1997) utilisation-focus. Essentially, utilisation-focused evaluation seeks to ensure that the conduct of an evaluation, and its results, are relevant to the questions asked by those who have a stake in the programme, and is useful in so far as it can effect positive change.

In any evaluation, there are many potential stakeholders with varying degrees of interest in the programme. To prevent use being stunted by the numbers of stakeholders, a key initial task for the utilisation-focused evaluator lies in undertaking a stakeholder analysis to identify the *primary intended users*, and then working with them to further identify the *primary intended uses*:

Evaluations must be focused in some way; focusing on intended use by intended users is the most useful way. (Patton, 1997, p.382)

The evaluator's job, then, in the early stages is to facilitate a working relationship with primary intended users to help them consider what kind of evaluation questions they want to ask, or reconsider the evaluation questions they already have in mind, and then determine what evaluation design would be appropriate.

Whilst working in partnership with local voluntary agencies, the INTEGRA Project was a Criminal Justice initiative, bid for, and operating under the auspices of, County Durham Probation Service (CDPS). Its main beneficiaries were CDPS probationers whose access to employment and/or training was additionally hindered by problematic alcohol and/or other drug use. As such, it was clear from the outset that the organisational primary intended user was CDPS.

The initial design characterised above, and the questions it was meant to address, were arrived at by my evaluation manager in consultation with CDPS management. In determining the questions and the design, both parties brought some 'preference' to their consultation.

CDPS and 'What Works'

As outlined in Chapter 3, and against a backdrop of 'Nothing Works' pessimism, political pressure on the Probation Service to demonstrate effectiveness has intensified over recent years (Chapman and Hough, 1998). Faced with such an intense *proving* backdrop, evaluative studies of Probation programmes have increasingly tended to adopt a methods focus from the positivist traditions. As such, experimental and quasi-experimental designs are the favoured route to providing best evidence (Nutley and Davies, 2000).

Academic Evaluation Research

In terms of the distinction between academic and service evaluators (see Chapter 4) I have suggested that the evaluation manager, with a background in academic evaluation research, and an openly proclaimed leaning towards positivism, is characterised by the notion of *academic evaluator*.

In essence then, the What Works political pressure experienced by CDPS, and the experience/preference/philosophy of the evaluation manager, coalesced to drive the initial evaluation design towards its experimental approach.

Theory and Suitable Measures of Impact

The underlying theory of the INTEGRA Project was that: "Unemployed offenders who find a job are less likely to re-offend than those who remain jobless". More specifically, INTEGRA was to target unemployed offenders with an identified problem associated with their alcohol and or other drug use. These were offenders

whose employability was thereby additionally hindered, and who were deemed to be even more at risk of social exclusion.

With an experimental backdrop to this evaluation now in place, the task was to take the theory into account in formulating appropriate evaluation questions. The key outcome to be addressed was "re-offending". How could this be measured to assess the impact of the INTEGRA Programme?

In the past, "reconviction" rates as a measure of Probation effectiveness have been used as a proxy measure for re-offending (Wilkinson and Morgan, 1995; Roshier, 1995; Oldfield, 1996; Richards, 1996). However, these present some disadvantages as outlined in Chapter 4, and as reviewed by, amongst others, Mair and Nee (1992) and Lloyd *et al*, 1994. Given such difficulties, there is interest in other possible measures of impact, such as attitudes, self-esteem, and social problems which might provide more relevant proof of 'What Works' (McGuire, 1995).

At the same time, direct use of drugs or alcohol could be measured, as could respondents' attitudes about their substance use. This approach was therefore adopted on pragmatic grounds.

European Social Fund Auditing

In addition to outcome indicators, the Project had a very real need to demonstrate its effective spend of £2.5 million from Europe. Failure to meet targets could result in funding being reclaimed, and as this had already been spent it would need to be found elsewhere within the organisation. The Project had set itself 'ambitious' output targets for the numbers of beneficiaries who would enter jobs or receive training. The potential for output anxiety within the Project was high. Outputs, then, required measurement and monitoring.

INITIAL EVALUATION QUESTIONS

Consequently, fundamental output evaluation questions focused on the numbers of service users who got jobs or undertook training courses.

Bearing in mind the Probation Service 'What Works' imperative, and particularly the advantages of those outcomes that do not suffer the same difficulties as 'reconviction', other evaluation questions related to the Project's impact on users':

- Attitudes to offending?
- Level of skills?
- Self-esteem?
- Attitudes to work?
- Alcohol and/or other drug use?

Additionally, the evaluation sought to address questions relating to the mean costs of outcomes, factors which might predict successful outputs, and service user/staff satisfaction with the Project.

This, then, was the original evaluation design. An experimental, longitudinal, and summative study to address the above questions. A Stakeholder Evaluation Management Group was established as a subgroup of the INTEGRA Management Committee to oversee and support the evaluation. This was the point in the process that I was appointed, under a PhD Studentship, to work as a Research Assistant under the supervision of the evaluation manager, the initial designer.

The plan was for me to administer an array of questionnaires to two cohorts of INTEGRA users, and re-interview them 12 months later. These interviews would take place in suitable probation offices across Durham County. Subsequent statistical analyses of the data would seek to answer the evaluation questions

INPUT OF INTEGRA STAFF AND THE 'SERVICE' EVALUATOR

As mentioned in my Introduction to this thesis, I came to this venture as a former social work practitioner/trainer/manager, working in a variety of agencies in the statutory and voluntary sectors, and mainly specialising in the drugs field. I had previously undertaken some small internal evaluations associated with some of my posts, and had experience of being externally evaluated. I had some limited understanding of evaluation. However, I also came to this endeavour with the notion, and motivation, that a commitment to good evaluation was a requirement of responsible caring.

As I was to realise, my understanding of evaluation was essentially characterised by Shadish and Epstein's (1987) notion of a *service evaluator*, with a preference for serving stakeholder needs, assisting programme improvement and decision-making, and formative evaluation approaches. I also brought, albeit unarticulated at that stage, a philosophical leaning towards constructivism – a point to which I will return a little later.

As outlined in Chapter 4, I became concerned at an early stage that the evaluation would do very little to assist the Project in its European Social Fund directive to be innovative. Project staff, too, had similar concerns. They were asking for something more formative, and something more qualitative:

QUALITATIVE AND FORMATIVE

After some discussion with my manager, and in addition to the initial quantitative and summative design, the evaluation also developed qualitative and formative characteristics. The qualitative components sought to capture the participants' experience of the Project as expressed in their own words. They were underpinned by a social justice commitment (House, 1991, 1999), and sought to enhance inclusion, dialogue, and deliberation. By so doing I believed that the evaluation might assume greater external validity – complementing my manager's rigorous empirical focus on the evaluation's internal validity. The qualitative components also sought to uncover propositions that might be suitable, resources and stakeholder approval permitting, for further systematic exploration, either quantitatively or qualitatively, as the Project evolved (Shaw, 1999).

The formative elements involved my developing a close working relationship with the Project and its staff, feeding back information and opinion wherever I believed it might be helpful (Stake, 1975).

Clearly, as described in Chapter 2, such an evaluative stance would raise objections from traditional evaluation theorists, like Campbell and Scriven, who teach that data collection, for the results to be valid and reliable, should be separate from the programme. They would also exhort that a close relationship between the evaluator and the programme could undermine the evaluator's neutrality and independence:

It's quite common for younger evaluators to 'go native', that is, psychologically join the staff of the program they are supposed to be evaluating and become advocates instead of evaluators. (Scriven, 1991, cited in Patton, 1997 p. 113)

Scriven draws a distinction between what he regards as real 'evaluation' and 'evaluation consultancy', which is what he sees many other evaluators, including Patton, as undertaking. For him real evaluation should include a systematic attempt to maintain

internal validity and exclude sources of likely bias, including contact with programme staff:

...'the experts' value depends on the validity of their evaluation, and that evaluation is compromised by social interaction...distancing has its price, but involvement risks the whole capital (Scriven, 1997, p. 488).

Indeed, this was just such a concern expressed by my evaluation manager. However, other than our discussions around the closeness of my stance, I felt it was also important for my manager to realise that the qualitative components were not intended to be a tokenistic gesture, bolted on to make her positivism look more 'realistic', or to provide a few 'juicy' quotes from probationers to enliven what might be read by some as another boring report.

To assist mine, my manager's, and others', understanding of what type of evaluation design this was, I constructed the amoeba metaphor in Chapter 4. The design, then, was simultaneously tight and loose, simultaneously quantitative and qualitative, and simultaneously summative and formative. In Patton's (1997) terms, it was 'situationally-responsive', and it was 'active-reactive-adaptive'. In my terms, it was 'organic', and, hopefully, capable of navigating the real world 'messy' conditions of the evaluation process. At its core, in addition to the experimental original design, also lay the Guiding Principles of the American Evaluation Association underpinning my stance in relation to evaluation.

It was around this point in the evaluation process, that my manager informed me I was now the "lead researcher". However, I was also a student and she was a senior lecturer, as well as the INTEGRA evaluation contract holder. As far as I was concerned, she was still my manager.

The development of this evaluation's design represents the first site of meaning generation within the INTEGRA evaluation. It represents a combination of theory, politics, research findings, practice preference, guiding principles, and philosophy, all coming together to create the baseline material, the evaluation approach and design, from which meanings could be constructed.

WHAT KIND OF PHILOSOPHY?

As Caracelli and Greene (1997) note, writing and reviewing from a US perspective, evaluators rarely present their work with an explicit statement of the philosophical or value frameworks that guided them. Without such information, we are left to ponder whether quantitative and qualitative methods were used within a single paradigm or multiple paradigms.

In line with the amoeba metaphor, this evaluation design was simultaneously positivist and constructivist – generally. Reflecting the power differentials that existed between us, as described on the previous page, I think it is fair to say that my brand of constructivism had a constant struggle to exert its voice to be heard equally alongside my manager's brand of positivism.

In Chapter 2, I summarised two poles of a paradigm spectrum as a way of introducing my exploration of the evaluation landscape as depicted by selected key figures, and illustrating the multifaceted philosophical terrain. By situating myself within this terrain I concluded that I have a leaning towards the constructivist end of the spectrum, but do not embrace relativism too tightly, except in my more sceptical moods. Whilst I do not recognise objective reality out there, I do recognise the existence of embedded social systems, and, with evaluation firmly in mind, this is no more apparent than when considering social policy and social programmes. I also share a realist's sense that fallible knowledge about these social systems is attainable. My constructivism, though, centres around the contextual construction of this knowledge, the importance of values and subjectivity, my unease around causation, and my uncertainty regarding transferability to other settings.

On the other hand, my former evaluation manager openly wears the badge of 'positivist'. In his critical analysis of UK contemporary evaluation research perspectives, as specifically applied to the evaluation of practice, and as mentioned in Chapter 2, Kazi (2000) notes the view of Outhwaite (1987) that there are at least twelve varieties of positivism and therefore:

...any single description of positivism as a paradigm will not be able to do full justice to all these variants. (Kazi, 2000, p. 758).

Given such complexity, I do not think any further consideration of my former manager's particular brand of positivism is required for

this thesis. Suffice to say, whilst the amoeba design for the INTEGRA evaluation successfully, I feel, accommodated our differing philosophical stances, this was not without some discomfort and disagreement. Nevertheless, I believe the additional effort and thought required to overcome some disagreements benefited this evaluation, as well as myself.

Towards the end of Chapter 2, I outlined my personal construct of evaluation. It has been interesting to reflect on this as I have proceeded through the various stages of this thesis. Whilst I feel that my fundamental philosophical stance, and associated evaluation construct, remains much the same, there are two theorists who have become somewhat more important to my sense of evaluation, Pawson and Tilley (1997).

In many ways, my use of Programme Logic Models to depict the INTEGRA Project at inception and close, taken together with my post-evaluation interviews with primary intended users, seem to represent one lap of Pawson and Tilley's Realist Evaluation Cycle. Certainly, programme theory, and the importance of process and context, as well as striving to construct a holistic snapshot of the INTEGRA Project, underpinned much of my stance in relation to methodology and consequent choices as they emerged. Does this mean I was embracing scientific realism more than I appreciated? No doubt I will continue to pursue this question as I participate in future evaluations.

What has been presented in this Chapter so far, describes a relatively tidy picture of the evaluation design process, and culminates in a relatively tidy metaphor. However, beyond design....

METHODOLOGY IS A 'MESSY' BUSINESS

As Patton (1997, p.118) notes:

The real world doesn't operate under textbook conditions. Effective evaluators learn to adapt to changed conditions. This requires situational responsiveness and strategic, contingency thinking – what I've come to call being active-reactive-adaptive in working with primary intended users

So why did probationers prove so difficult to recruit? The design seemed comprehensive and yet flexible. Mixing methods appeared

to be a suitable strategy to meet the competing demands from the summative 'What Works' heritage, and the ESF directive that the Project innovate, and thereby require something more formative to assist in this. And at the level of paradigms, my manager and myself had, generally, found a way to accommodate our differing worldviews, not least through an apparent shared commitment to utilisation-focus.

Nevertheless, a quantitative core methodology, a 'hard-to-reach' user group, and no *tangible* incentive: so goes the conundrum at the heart of this evaluation. Just how do you go about recruiting an adequate sample ('getting the numbers'), and how do you do it ethically?

At the outset of my contacts with INTEGRA staff, in addition to their wanting something more qualitative and formative, they also expressed concerns that I would not get to see sufficient numbers of their 'hard to reach' clients to satisfy the statistical requirements of the evaluation core design.

During the pilot phase of the data-gathering, twenty five office-based appointments were made. One person turned up and there were no cancellations.

WHY SO 'HARD TO REACH'?

In Chapter 4, I outlined how drug users in treatment tend to be available for 'research' approaches (Parker and Kirby, 1996; Bloor *et al*, 1997). Principally, there is usually a 'recruiting' point, a treatment centre. However, the INTEGRA Project worked on a one-to-one outreach basis in CDPS premises. Presumably too, but not always, drug users in treatment may be more motivated. INTEGRA clients, on the other hand, given their 'no-choice' referral for initial assessment, might include in their midst a significant number of unmotivated PRECONTEMPLATORS (Prochaska and DiClemente, 1984). Additionally, set appointment times might not fit comfortably with aspects of problematic drug users' variously chaotic lifestyles. Drug users, too, might feel they had something to risk in meeting a researcher:

Those who live drugs-crime lifestyles are not just paranoid. People really are out to get them. (Brain *et al*, 1998).

And finally, there was no tangible incentive, no payment. As Hughes (1999) notes, a small *material incentive* is a practical way to encourage people to participate in research.

ETHICAL DECISION-MAKING

To overcome such difficulties, it was suggested by CDPS that we make evaluation contacts 'reporting contacts' under National Standards (Home Office, 2000). My manager agreed with this approach, after all, the core quantitative design for this evaluation was at risk: no numbers, no credible analysis. As such, any 'good' that could come from the study might be lost.

My manager felt no intuitive ethical concerns regarding such a strategy. But I did have such concerns. Adopting Newman and Brown's (1996) framework for ethical decision-making (see Chapter 5), I attempted to systematically translate my intuitive disquiet into cognitions. Reflecting a social justice commitment to inclusion (House, 1999) I recruited a small group of probationers to discuss the issues, I also consulted widely with other researchers, and I 'tested what it was like' to undertake an evaluation contact as a 'reporting contact'. I arrived at the conclusion that such an approach was, indeed, unethical.

Given that 'avoiding harm' to respondents lies at the top of my personal ethical construct, I felt I could not proceed with a methodology that may place respondents in jeopardy of further punishment. As outlined earlier, there are sound reasons why drug users may not wish to meet with a researcher. Likewise, even if I adopted a position that the evaluation contact was 'reportable', but that this did not imply that any 'sensitive' evaluation information had to be given by the respondent, I would be ignoring the impact drug use could have on their cognitive function. It is all well and good my appreciating such a subtle difference, but variously chaotic drug users may well feel they 'have to' provide information because it is a reporting contact. Indeed, in Chapter 5, I describe how my 'test run' of 'reporting evaluation contacts' resulted in the respondent, who was not a chaotic user at that stage, believing he had to answer my questions because it was part of Probation. Likewise, his mother believed the same.

By asking people to convey information about their drug use with the Maudsley Addiction Profile (Marsden *et al.*, 1998), and, particularly at T2, by asking about their current benefit fraud and other offending, I was placing them at possible risk if exposed.

Given that such information, regardless of my assurances of confidentiality (Fitzgerald and Hamilton, 1996), could be open to court scrutiny under certain conditions, I believed the CDPS offer of 'reporting contacts' undermined the rights of respondents to respect and autonomy, and placed them at risk of possible harm.

Whilst such a stance might have jeopardised the quantitative core of the evaluation, I felt that this could not justify 'infringing' respondents' rights. The notion of 'infringing' a right refers to a justified action overriding a right (Beauchamp and Childress, 2001), and would be, under certain justifiable circumstances, the course of action undertaken by those evaluators with a broadly utilitarian stance. This, indeed, I believe informs much of my ethical position - to make a decision based on the particulars and *consequences* of a specific situation rather than search for a rule applying to all similar situations.

Reflecting on this particular ethical conundrum at the core of the INTEGRA evaluation, and given my utilitarian stance of consequentialism, I believe my constructivist hesitation played some part in my ethical decision-making. Given my views regarding reality and knowledge, I believe myself to be less inclined towards any 'justified' infringement of respondents' rights, and more inclined to classify any potential harm as 'undue' than, say, some positivist empiricists 'rigorously' searching for, what I believe to be, the naïve realism of 'empirical truth', and holding, what I again believe to be, naïve expectations of 'the good' that could come from the search. Who cares if the quantitative core of the design was compromised by a commitment to avoiding possible respondent harm?

Ethical Compromises and Contactability Bias

As described in Chapters 4 and 5, my ethical decision regarding 'reporting' contacts necessitated an adjustment to the methodology. Home visits became my preference, not least given the degree of discomfort experienced by some respondents, and members of my recruited probationer group, in discussing drug use and other 'sensitive' issues within a probation office.

In arranging home visits, however, I did not spell out the notion of 'informed consent', leaving this until I met face-to-face with the person. I also decided on an 'opt-out' rather than 'opt-in' approach. Whilst I was, therefore, more pushy in an attempt to 'get the numbers', I felt ethically justified in doing this. There seemed little chance of doing 'harm' to the recipients of my letter, and any subsequent interview would still only proceed with their signed

consent. Whatever the ethical price of proceeding this way, this had to be balanced against the ethical price of not getting to see sufficient numbers to validate the evaluation, thereby undermining it, and undermining the possible 'good' that could come from its use.

My constructivist hesitation, then, was not so entrenched that it did not permit some ethical compromise. But I could not compromise regarding possible harm to respondents. In this, I believe my social work roots, and particularly in specialising within the drugs field, had some bearing. In Chapter 2, I described what I felt to be the unavoidable web of advocacy inherent in evaluation practice:

Evaluation inherently involves advocacy, so the important question becomes advocacy for whom. The most defensible answer to this question is that evaluation should advocate for the interests of program participants. (Greene, 1995, p.1)

Whilst others would not agree, including Patton (1997) who, too neatly I feel, sees advocacy as a matter for discussion with primary intended users, I have to admit that advocacy is a component of my understanding of evaluation practice. In my social work roles, and since the late 1970s, I have spent much time advocating the rights and welfare of drug users, offenders or otherwise. It is not something I am likely, or indeed would want, to forget very easily. As House (1980, p. 117) puts it: "Truth is more important than beauty. And justice more important than either."

However, it is clear that the INTEGRA sample contained systematic bias as a consequence of my ethical position. As described in Chapter 6, evidence of contactability bias was found in the sample which meant that people who were more likely to have used illegal drugs in the past, those who had more criminal acquaintances and those who were less positively disposed towards the Probation Service were under-represented in this evaluation. By promoting respondent autonomy in the way I did, it seems I got to see the less problematic end of the INTEGRA client spectrum.

As a way of exploring the notion of contactability bias further, I decided, and with stakeholder approval, to explore respondents' motivation to participate in the evaluation, using themes identified by Hughes (1999). Whilst I had already persuaded CDPS to pay respondents for T2 interviews in an attempt to minimise sample attrition over time, my hypothesis that payment would be viewed as an important motivator by respondents was not validated by their

responses. However, given the pressure they may have felt to provide socially desirable responses in my presence, I still suspect that had a £10 payment been offered for T1 appointments the issue of recruiting a suitable sample size would not have been so problematic.

ADDITIONS AND THE PSEUDOPODIC FRINGE

The exploration of respondent motivation to participate was one of the additions to this evaluation brought by a commitment to be 'situationally responsive' (Patton, 1997). It represents the activity in my amoeba metaphor which lies at the pseudopodic fringe, the systematic exploration of issues as they emerge and are also felt by the primary intended user to be worthy of further study. Other contemplated additions to the design were:

- Case Manager survey;
- Greater semi-structured exploration with probationers at T2, particularly exploring the impact of the informal economy which had arisen as an issue in an unsystematic way during T1 Balance Sheet discussions;
- Partnership issues, particularly the relationship between Astep and E Team, and the discrepancy between their operating principles of abstinence and harm-minimisation respectively;
- Semi-structured 'exit' interviews with all staff.

Reflecting on why I believed such additions were relevant:

1/. Contactability bias and respondent motivation was explored as the reliability of outcome data would be an issue in terms of a utilisation-focus. If the data were seen to be flawed in some way, our findings would be less likely to be used by policy-makers. In Chapter 2, I noted the work of Weiss and Bucuvalas (1980) in linking policy research to evaluation. Their research suggests that policy-makers employ filters when looking at research findings. Weiss categorises these implicit filters as 'Truth Tests', which include the perception made by policy-makers of the *research quality*, and 'Utility Tests', whether the research gives explicit guidance for feasible reform, or whether more fundamental change, with more consequent uncertainty and risk, is suggested

2/. The Case Managers survey was seen as a way of reflecting the Probation context in which the Project was sited. I believed this was an important consideration in developing a more holistic picture of

the Programme Theory, and what constraints (or determinants) may exist in the wider context, either by not supporting the Project premise, or by not viewing the Project positively for other reasons. If problems had been found, then, working formatively, it would allow the Project to take some action, as it did in relation to some reported uncertainty about its working methods.

3/. Client participation in the informal economy, and other possible disincentives to formal job-seeking could have an impact on Project outputs and outcomes, and the development of Programme Theory. Client satisfaction with the service, and their attitudes towards drugs, work, and offending were questions which the initial evaluation design was committed to addressing. Exploring these in a qualitative way too, and comparing to the views of Case Managers and Project staff regarding the same issues would be useful as a source of varying perspectives and constructions in developing the Programme theory (Shaw, 1999; Guba and Lincoln, 1989.)

I should note here, that I am not suggesting that such *triangulation* (Denzin, 1989) somehow improves the validity of the data; that by combining quantitative and qualitative exploration of the same phenomenon we somehow get a: "geographical 'fix' on the true 'location' of evidence" (Shaw, 1999, p. 186). Such naïve realism does not sit comfortably with my constructivist hesitation.

I should also note here, that the notion of the 'informal economy' is more complex than I have outlined here. Suffice to say for the purposes of this thesis, rather than thinking of one all-encompassing informal economy, it may be more useful to proceed with the idea of several economies (Davies, 1997; Shapland, 1997), and a blurring of the continuum between the formal and the informal. Such economies could be locality, race, and gender dependent, may involve crime, drugs, prostitution, 'fiddle work' as their nucleus, and they may well be interacting. For the purposes of this study, I have restricted my exploration of respondents' participation in the informal economy to that notion of 'working for reward without declaring such income for tax or welfare benefit purposes'.

4/. Partnership issues. In Chapter 4, I described how I became aware within the first few months of my involvement in this evaluation of problems between the operational aspects of the E Team and Astep. There was also a discrepancy between their operating principles that might have some bearing: the E Team worked on the principle of harm minimisation in relation to drug use,

whereas Astep expected probationers to be drug-free when referred to them. Astep believed that their drug-free expectation was shared by employers who had concerns regarding Health and Safety in the workplace.

The gap between worthy aspirations in policy and the reality of inter-professional relationships at practitioner level is well documented (Gifford, 1986; Lipsky, 1980; Nellis, 1995), and more relevantly for this evaluation, Rumgay and Cowan (1998) who report the problems of a number of Probation/Voluntary Sector partnerships working with 'substance misusing offenders'. Of particular importance in Rumgay and Cowan's view was a clash of cultures, with the Probation Service control and enforcement imperative being seen as problematic for some voluntary sector partners. Whilst withdrawal of cash in the event of a partnership breakdown was not seen as life-threatening for smaller voluntary agencies with a culture of entrepreneurship, it was still indicative of a significant imbalance in the power distribution. More recently, the evaluation of three Pathfinder Drug Treatment and Testing Order schemes (DTTOs) (Turnbull *et al*, 2000) noted:

Achieving effective inter-agency working is perhaps the single most important factor to address in establishing programmes. All three pilot sites encountered quite serious "teething problems"; and many of these remained unresolved. By the end of the pilot only the Croydon team had addressed conflicts and disputes sufficiently thoroughly to be operating as an effective team whose whole was more than the sum of its parts. (p. 82)

The problems identified were as a consequence of differences in working styles, traditions and values. Turnbull *et al* conclude:

We think they are likely to be widespread when DTTOs are rolled out nationally. (Turnbull *et al*, 2000, p. 82)

Whilst this DTTO final evaluation report was not available for INTEGRA consumption, being published towards the end of the Project, the interim findings (Turnbull, 1999) were available, and, indeed, were made available to Project management by me.

Given the centrality of the E Team/Astep relationship to the operation of the Project, and to it achieving its targets (see

Programme Logic Model, Chapter 3), I believed systematic exploration of the difficulties being experienced in the partnership might allow the Project to address these formatively and improve its functioning. However, despite a number of attempts to have this additional work sanctioned by the Probation-chaired Project Management Group, it never came to fruition. As such, the issue of partnership functioning was introduced into the staff exit interviews.

I believe my problems in trying to systematically access the E Team/Astep partnership may well have been political. I have no data to back this up, but a central concern throughout for the INTEGRA Project was the naivety of the original work-related outputs, and the consequent output anxiety for all involved – not least given the ESF retrospective funding procedures. Should the 'faeces hit the fan', someone would have to pay. Is this likely to be the more powerful Probation Service arm of the partnership?

5/. Exit Interviews with INTEGRA STAFF were essential I believe in understanding how the Programme, and its theory, was experienced by all staff. This, again I believed, would add greatly to the holistic picture of the Project.

Methodology represents the second site of meaning generation within the INTEGRA evaluation. A significant contribution here has been the impact of my ethical decision-making, and associated philosophical stance, in recruiting a suitable sample size, and, unfortunately but perhaps inevitably, in establishing systematic bias in the data. Likewise, significant contributions have emerged from adopting a situationally-responsive contingent stance in relation to the Project, and in developing further areas for systematic study. In essence, the evaluation evolved alongside Project operation and experience.

Generally, these further areas have been driven by the Programme theory, and once again I am left wondering how closely I actually embraced Pawson and Tilley (1997) throughout?

WHAT DID INTEGRA TELL US THAT IS USEFUL?

This is the third site of meaning generation within this evaluation. It is the point at which stakeholders, and others, react to the information generated, and the representational forms chosen to communicate that, to construct their own meanings. So what did INTEGRA tell us that could be useful?

With reference to the Programme Logic Model at Project close in Chapter 6:

- Offenders who find a job may be less likely to reoffend, but the nature of the job matters, and particularly in financial terms. Preferably it should pay more than people receive on welfare benefits PLUS whatever they can make elsewhere in some aspect the informal economy.
- Training and qualifications could enhance employability but many clients may not be interested.
- Overcoming alcohol and drug difficulties may improve employability, and may have a direct bearing on reoffending, but only if effective therapeutic interventions are offered.
- Employers are seen as unsympathetic and many offenders may prefer to jobseek independently so that they can deceive employers about the true extent of their offending or drug use.
- Ensure that targets are achievable, and that there is sufficient planning time to establish this, otherwise output anxiety may infect Programme process and performance.
- The Probation Service may not be the best place to site such a Project as INTEGRA, and could pose a block to innovation.

Firstly, it should be noted that the data which contributed most to the elucidation of the Programme theory emanated from the broad constructivist thrust of my qualitative approaches and analysis, and from the close working relationship I forged with the INTEGRA Project in order to permit a utilisation-focus.

My constructivist hesitation, however, now makes me somewhat uncertain about generalising such findings to other settings. Certainly, the rural make-up of Durham County should be borne in mind when considering the logistical problems experienced by this Project, notably travelling time. Likewise, being locally restricted to one-to-one outreach in CDPS premises also inhibited the potential for innovation and other forms of intervention, such as groupwork, and peer group mentoring. And finally, the sample make-up here consists of: all white British, mostly young, mostly male, and mostly with low or no skills.

Primary Intended User Feedback Regarding Use

In Chapter 7, I discussed the primary intended user response to the evaluation and its final report.

What was immediately clear was that late publication of the report had a negative impact on its capacity to be maximally useful locally. However, given the formative stance I adopted in working with the Project, information and opinion was fed back to the staff wherever it might be useful. As such, meaning was generated at the various points of maximal use, and helped develop CDPS understanding of how best to offer a County-wide service from within the Probation Service for drug-using offenders. This knowledge was to prove useful in CDPS implementing DTTOs. This became the exit strategy for the Project – the *primary use*, if albeit not *intended* at Project outset.

That said, CDPS felt that the evaluation report was still useful as an educational tool for some of its more clinical DTTO partners, and the data generated by the evaluation itself would have future use as a baseline from which to undertake future analyses.

The key to use of the evaluation, I feel, lies in the close communication and relationship that was established between myself and those I identified within Project/Probation management as the primary intended users. As Patton (1997) notes, to not enter the 'fray' and get involved with projects is a major diversion from a utilisation-focus. And in getting involved, it's the people, not the organisation, that matters the most.

Whilst Patton also notes that another major diversion away from a utilisation-focus is to: "assume that the funders of the evaluation are the primary intended users..." (Patton, 1997, p. 57), I have been somewhat surprised by the lack of European Social Fund interest in this evaluation's findings.

May Pettigrew (2002), the President of the UK Evaluation Society, writes:

My perceptions now are of a [European] Commission in which evaluation cultures are strong and increasing. Evaluation is mandated and ubiquitous. The quality or commitment to the use of findings and feedback is much more variable.

I found it educational to browse the Evaluation website of UK ESF. Whilst the ESF evaluation team occasionally commissions independent evaluations to follow-up 'Leavers', they primarily undertake evaluations themselves based on the 'closure forms' of projects, based on the paper submitted at project end. I find it

personally fascinating that so much data in electronic form (databases, spreadsheets, or whatever) are left sitting on computers, presumably bought with ESF monies, up and down the country.

Such a laissez-faire attitude to potential knowledge generation can be compared to the commitment shown by the Department of Health (DOH) under its research governance framework. This sets standards, details the responsibilities of the key people involved in research, outlines the delivery systems and describes local and national monitoring systems. The DOH website also offers The Research Findings Electronic Register (ReFeR), which is a freely available database, providing 'prompt sight' of quality assured information on research findings that emerge from completed projects funded by the DOH, including the NHS Executive.

Perhaps the European Social Fund could benefit from a browse of the DOH site themselves?

BUT WHAT OF EVALUATION ITSELF?

What might this INTEGRA evaluation have to say which could be useful to the art of evaluation itself?

That evaluation is an art is not a controversial idea to me. But whilst I appreciate evaluation as an art, I am also aware that it requires discipline and focus if it is to produce findings that are robust, reliable, and valid, and if it is to encourage use of these. As its range of creative approaches and tools widens, evaluation must continually learn its lessons and retain a commitment to discipline and focus if it is to fulfil its promise.

As with my reporting of the programme-specific findings from the INTEGRA evaluation earlier, my constructivist hesitation makes me somewhat uncertain about generalising broader evaluation lessons drawn from my experiences of this one study. Nevertheless, there are a number of points I would like to make which may be useful for others:

1/. Philosophy matters. This does not mean that I wish to see a return to the noise of the paradigms debate outlined in Chapter 2. However, I believe it behoves each of us, if we strive for any sense of evaluation professionalism, to explore and articulate our philosophical perspectives and be alert to how these impact on all aspects of our work. This is something I have attempted to do

throughout this thesis. Likewise, as illustrated in Chapter 7, I believe we need to be alert, as far as possible, to the philosophical perspectives of our 'primary intended users' and how these might impact on our chosen forms of representation and the meanings our users construct from these.

2/. Develop a multi-method toolkit. Philosophical awareness does not imply some notion of paradigm purity and a consequent narrowing of our design and methodological repertoire. As outlined at the end of Chapter 2, and although underpinned by a broad sense of constructivism, my personal construct of evaluation embraces a number of theorists. Even if I do not necessarily agree with their theoretical justifications, or share their philosophical perspectives, there are design and methodological practices in the work of each of these theorists that I would endorse. Good practice, I believe, is eclectic and informed by theory, not puristically derived from it. Just as a narrow focus on experimental and quasi-experimental approaches proved insufficient to the task in evaluation's early years, so today a narrow focus on any one other approach is equally insufficient to the complexities of the field. A credible evaluation construct needs a place for multiple approaches, and must be clear about the contingencies which guide the selection of one over another when programme circumstance, stakeholder requirements, and resources force that choice.

3/. Get a good grounding in applied ethical thinking. My motivation for including Chapter 5 in this thesis was driven by the conundrum at the heart of the INTEGRA evaluation: '*a quantitative core methodology, a 'hard-to-reach' user group, and no tangible incentive: just how do you go about recruiting an adequate sample ('getting the numbers'), and how do you do it ethically?*' In my approach to this conundrum I found Newman and Brown's (1996) framework for ethical decision-making invaluable. I came to realise that evaluators, from the very start of an evaluation to the very end, encounter various situations which demand a grounding in applied ethical thinking. I also came to realise that this goes far beyond the, at best, ethical minimalism of a reliance on prescribed ethical codes or the 'outset approval' of some Ethics Committee. Striving to be ethical, of course, does not mean that we will always make the best decisions or the best choices. But it does mean that we will keep trying.

4/. Pictures can be powerful tools. I now believe that the process of creating programme logic models is an important component of any evaluation. However, unlike my use in this thesis, this, I feel, is

something I should have undertaken with the INTEGRA Project Management Group and repeated on a regular basis as a routine way of monitoring programme process and exploring programme theory. That way, for instance, the partnership difficulties between the E Team and Astep would have received a regular and very visual consideration. Likewise, I believe the use of metaphor to be particularly important in encouraging stakeholder understanding of, and participation in, the evaluation process. Certainly, INTEGRA staff response to the amoeba metaphor presented in this thesis was very positive.

5/. Wherever feasible, revisit our clients. I believe that we need to follow our reports back to our clients and talk about them. Certainly, the exercise undertaken in the INTEGRA evaluation of seeking primary intended user response to the evaluation report proved informative. However, wherever feasible, I also believe we should recruit someone independent to undertake this work. Given my closeness to the INTEGRA Project, a closeness I feel was necessary to sustain a utilisation-focus, I believe it may have been more difficult for primary intended users to be forthright about negative aspects in my presence. An independent enquirer may have generated a more rounded response.

And that is all I feel comfortable saying....

APPENDIX A
Information Sheet
+
Consent Forms



Centre for Applied Social Studies
15 Old Elvet, Durham DH1 3HL
tel 0191 374 4731

INTEGRA Project Evaluation INFORMATION SHEET

Research shows that finding a job helps many offenders achieve a crime-free life. It is important, therefore, that all projects which try to help them find work are examined and improved.

The INTEGRA Project aims to help offenders in County Durham overcome problems in their lives which make it even more difficult for them to find work. In particular, INTEGRA aims to help people deal with any problems arising from their alcohol and/or other drug use, and then take steps to get work, training, or further education. Although it works in partnership with local voluntary agencies, it is a Probation Service project which works with offenders on community sentences or following their release from prison.

The INTEGRA Project is funded for 2 years by the European Social Fund.

The Centre for Applied Social Studies (CASS) has been asked to research the Project to find out how well it works. As part of this research, CASS will be interviewing two groups of 50 INTEGRA users, as well as a control group of 50 standard probationers. Each group will be interviewed again 12 months later to see whether there have been any benefits from INTEGRA attendance.

These interviews are strictly confidential and will give people the chance to talk about the help they receive/have received from INTEGRA, and whether or not they found that help useful.

If you would like more information about this research, then please feel free to contact the researcher:

Paul Burlison at the above address or on direct line **0191 374 4731**

or the research supervisor, Dr Justine Schneider, at the above address.

Thank you.



Centre for Applied Social Studies

**INTEGRA Project Evaluation
CONSENT FORM TO TAKE PART IN RESEARCH**

- I understand that this research is designed to evaluate the work of the Durham County Probation Service INTEGRA Project.
- I have had the research explained to me and have been given an information sheet.
- I understand that I can withdraw my consent at any stage without giving reasons, and without prejudice to me.
- I understand that my participation in this research will involve two interviews, each about one hour in length (one now, and a further interview in 12 months).
- I consent to take part in this research project, and additionally consent to the researchers contacting me again in 12 months to arrange the second interview.

NAME _____

SIGNATURE _____

ID NUMBER:

--	--	--	--	--

DATE:

--	--	--	--	--	--

INTERVIEWED BY:

--



University
of Durham

Centre for Applied Social Studies

INTEGRA Project Evaluation
CONSENT FORM TO TAKE PART IN RESEARCH

- I understand that this research is designed to evaluate the work of the Durham County Probation Service INTEGRA Project.
- I have had the research explained to me and have been given an information sheet.
- I understand that I can withdraw my consent at any stage without giving reasons, and without prejudice to me.
- I understand that my participation in this research concludes with this interview, having already been interviewed 12 months ago.
- I consent to take part in this research.

NAME _____

SIGNATURE _____

ID NUMBER:

--	--	--	--	--	--

DATE:

--	--	--	--	--	--	--

INTERVIEWED BY:

--



University
of Durham

Centre for Applied Social Studies

INTEGRA Project Evaluation
CONSENT FORM TO DISCLOSE INFORMATION

- I understand that some of my responses in this interview have created concerns for my welfare, or the welfare of others, in the researcher.
- I have had these concerns explained to me.
- I consent to the researcher informing the below named about these concerns.

NAME _____

ROLE _____

NAME _____

SIGNATURE _____

ID NUMBER:

--	--	--	--	--	--

DATE:

--	--	--	--	--	--	--	--

INTERVIEWED BY:

--

APPENDIX B

Client T1 Balance Sheet

**INTEGRA EVALUATION
BALANCE SHEET**

ID Number:

Date

Interview Length (mins)

**YOUR VIEWS ON THE POSITIVES AND NEGATIVES
OF YOUR CONTACT WITH INTEGRA**

POSITIVES	NEGATIVES

Would any of your responses to these questions have been different had you been interviewed _____?

YES

NO

PLEASE WRITE ANY COMMENTS OVERLEAF

APPENDIX C
Client Interview T2
+
Motivation to Participate

INTEGRA EVALUATION CLIENT "EXIT" INTERVIEW SUPPLEMENT

ID NUMBER

--	--	--	--	--	--

DATE:

--	--	--	--	--	--

INTERVIEWED BY:

--

TAPÉ DONE?:

--

INTERVIEW LENGTH:

--	--	--

LISTED AS OUTPUT?

--

--

1. UNEMPLOYED OFFENDERS WHO FIND A JOB ARE MUCH LESS LIKELY TO REOFFEND THAN THOSE WHO REMAIN JOBLESS. COMMENTS?
2. IT DOESN'T MATTER WHAT SORT OF JOB THIS IS. COMMENTS?
3. IT'S UNREALISTIC TO EXPECT OFFENDERS TO ACCEPT A WAGE WHICH IS NOT MORE THAN THEY RECEIVE ON BENEFIT. COMMENTS? WHAT ABOUT INFORMAL ECONOMY?
4. TELL ME ABOUT THE WORK/TRAINING HELP YOU RECEIVED FROM THE PROJECT & WHAT HAS HAPPENED SINCE.
5. HOW HAS CONTACT WITH THE PROJECT AFFECTED YOUR DRUG/ALCOHOL USE? (INCLUDING ATTITUDES)
6. HOW HAS CONTACT WITH THE PROJECT AFFECTED YOUR OFFENDING? (INCLUDING ATTITUDES)

7. WHENEVER RESEARCHERS CONSIDER WHY PEOPLE AGREE TO PARTICIPATE IN RESEARCH, IT IS GENERALLY THE VIEWS OF THE RESEARCHERS THAT ARE CONSIDERED – BUT WHY DO YOU THINK YOU AGREED TO PARTICIPATE BY BEING INTERVIEWED BY ME?

FOR INSTANCE:

DID YOU FEEL IT WAS SOMETHING YOU HAD TO DO, YOU HAD NO CHOICE?

WHAT ELSE?

- Interesting, relevant, and worthwhile research;
- The timing of the fieldwork and your priorities;
- The benefits of the research to you;
- Research benefits to people in general;
- The impact of others on participation;
- Your first impressions of the researcher;
- The importance of confidentiality;
- The use of tangible rewards and incentives.
- ANYTHING ELSE?
- WHAT DO YOU CONSIDER WAS THE SINGLE MOST IMPORTANT REASON?

• ANY OTHER COMMENTS? (OVERLEAF)

APPENDIX D
Staff Exit Interviews

INTEGRA STAFF EXIT INTERVIEW

With:

Date:

Role:

1. Unemployed offenders who find a job are less likely to reoffend than those who remain jobless. Comments?
2. It doesn't matter what sort of job this is. Comments?
3. It's unrealistic to expect offenders to accept a wage which is NOT more than they receive on benefit. Comments?
4. In your experience, how have Probation Officers responded to your work? As far as you can say, is this response typical of their overall response to the work of INTEGRA?
5. In your experience, how have clients responded to your work? As far as you can say, is this response typical of their overall response to the work of INTEGRA?
6. From your perspective, how have the partnership arrangements for INTEGRA worked out in practice?
7. From your perspective, how have the transnational partnership arrangements for INTEGRA added value?
8. How helpful have been the joint referral criteria of alcohol/drug problems AND unemployment?
9. How helpful has been the automatic referral process?
10. The INTEGRA Project was charged with being innovative. How far would you say this has been achieved? What would you consider to be the single most important aspect of this innovation? What would you consider to be the single most important block to innovation potential?
11. Describe and assess INTEGRA response to issues of:
 - a/. YOUR PERSONAL SUPPORT AND DEVELOPMENT
 - b/. YOUR PROFESSIONAL SUPPORT AND DEVELOPMENT
 - c/. OVERALL TEAM SUPPORT AND DEVELOPMENT

INTEGRA POSITIVES

INTEGRA NEGATIVES

APPENDIX E
Case Manager Survey

**DURHAM UNIVERSITY
CENTRE FOR APPLIED SOCIAL STUDIES
INTEGRA EVALUATION**

CASE MANAGER QUESTIONNAIRE

What is this survey?

This is a survey of your views about offender employment interventions in general, and about the County Durham Probation Service (CDPS) INTEGRA Project in particular.

As a Case Manager, you are a major stakeholder in any initiative undertaken by CDPS, and any evaluation of the kind we are currently undertaking would be incomplete without your feedback.

Who will see my answers?

The information you give is confidential. A report will be compiled describing aggregated questionnaire results. Individual respondents will not be identified. There is a brief "professional experience" section at the end to enable the introduction of "experience" variables into our analysis.

However, as we may seek to investigate some responses in more detail, you may identify yourself in the section provided at the end should you wish.

We would be grateful if you could complete this questionnaire and return it (unfolded and not stapled) to Paul Burlison at the Centre for Applied Social Studies in the A4 SAE attached. Paul is happy to discuss the evaluation if you would like further information (tel. 0191 374 4731).

THANK YOU FOR YOUR ASSISTANCE

Q1 Firstly, please indicate with a cross if one or more of your clients has had contact with INTEGRA:

YES

NO

If yes, how many?

--	--

se read each statement and place a cross in the box which best describes your
 ement or otherwise. (As questionnaires can sometimes raise more questions and qualifications
 answers, should you wish to expand on any of your responses then please do so on the comments page
 hed to the rear.)

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
employed offenders who find a job are less likely to reoffend than those who remain jobless.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
It doesn't matter what sort of job this is.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
It is unrealistic to expect offenders to accept a wage which is NOT more than they receive on benefit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Training schemes help people gain the skills they need to secure work.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Probation Service has a key role to play in helping offenders find employment or training.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
The Probation Service's role should be one of assessment and then referral to a specialist employment agency.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Probation Service is well-served by specialist employment agencies.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Arrival of the INTEGRA Project has confused referral process to employment agencies.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Probation Service has a key role to play in helping offenders deal with alcohol/drug problems.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Role should be one of assessment and then referral to a specialist alcohol/drug agency.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Area is well-served by specialist alcohol/drug agencies.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Arrival of the INTEGRA Project has confused referral process to alcohol/drug agencies.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Not clear about the objectives of the INTEGRA Project.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Not clear about the working methods of the INTEGRA Project.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

INTEGRA Project workers possess the skills to carry out their work effectively.

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
satisfied with the client feedback from the 1/2 drug phase of the INTEGRA Project.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
satisfied with the client feedback from the treatment phase (Astep) of the INTEGRA Project.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
INTEGRA Project makes a valuable contribution to CDPS aim of reducing reoffending.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

NOW FOR SOME INFORMATION ABOUT YOU:

Please list any professional qualifications you hold:

Please state how many years working for the Probation Service:

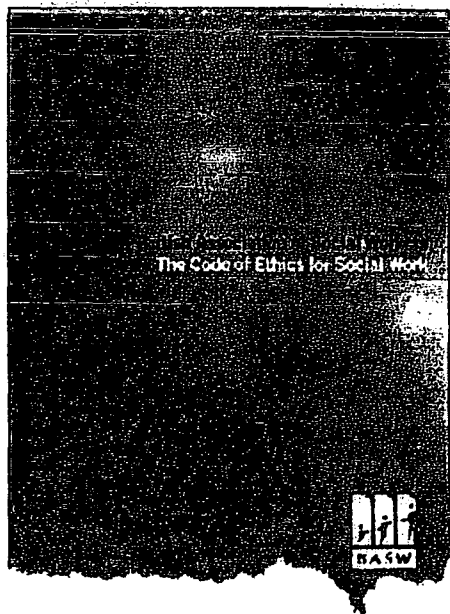
How many years working for CDPS?

Please state current team within CDPS:

Name and Contact Number
(Please provide if you wish to be identified)

1
ADDITIONAL COMMENTS?

APPENDIX F
BASW
Code of Ethics for Research



The Code of Ethics for Social Work

4.4.4 Research

In applying the general provisions of this Code, social workers engaged in research will observe the following specific ethical responsibilities.

- a. At all stages of the research process, from inception and resourcing through design and investigation to dissemination, social work researchers have a duty to maintain an active, personal and disciplined ethical awareness and to take practical and moral responsibility for their work.
- b. The aims and process of social work research, including choice of methodology, and the use made of findings, will be congruent with the social work values of respect for human dignity and worth and commitment to social justice. Social work researchers will therefore:
 - Predicate their work on the perspectives and lived experience of the research subject except where this is not appropriate;
 - Seek to ensure that the research in which they are engaged contributes to empowering service users, to promoting their welfare and to improving their access to economic and social resources;

- Seek to work together with disempowered groups, individuals and communities to devise, articulate and achieve research agendas which respect fundamental human rights and aim towards social justice;
 - Retain a primary concern for the welfare of research subjects and actively protect them from harm, particularly those who are disadvantaged, vulnerable or oppressed or have exceptional needs;
 - Consider and set out clearly how they would deal with the ascertainable consequences of proposed research activity for service users, in order to ensure that their legitimate interests are not unwarrantably compromised or prejudiced by the proposed investigation;
 - Not use procedures involving concealment except where no alternative strategy is feasible, where no harm to the research subject can be foreseen and where the greater good is self-evidently served.
- c. In accordance with their duty of competence, social work researchers will, in their chosen methodology and in every other aspect of their research, ensure that they are technically competent to carry out the particular investigation to a high standard. Where research is carried out primarily as an educational or instructional tool, this responsibility also falls on the student's supervisor.
- d. In accordance with their duty of integrity, social work researchers have a duty to:
- deal openly and fairly with every participant in the research process, including participants, service users, colleagues, funders and employers;
 - inform every participant of all features of the research which might be expected to influence willingness to participate, especially but not exclusively when access to services may be, or be perceived to be, affected by or dependent on participation;
 - in all cases respect participants' absolute right to decline to take part in or to withdraw

from the research programme, with special attention to situations in which the researcher is in any way in authority over the participant;

- ensure that subjects' participation in a programme is based on freely given, informed and acknowledged consent, secured through the use of language or other appropriate means of communication readily comprehensible to the research subject, conveying an adequate explanation of the purpose of the research and the procedures to be followed;
 - seek to exclude from their work any unacknowledged bias;
 - report findings accurately, avoiding distortion whether by omission or otherwise, including any findings which reflect unfavourably on any influential body or research sponsor, on the researcher's own interests or on prevailing wisdom or orthodox opinion;
 - seek to ensure that their findings are not misused or misrepresented;
 - acknowledge when publishing findings the part played by all participants and never take credit for the work of others.
- e. In accordance with their duty of confidentiality, social work researchers will respect and maintain the confidentiality of all data or information produced in the course of their research, except as agreed in advance with participants (including research subjects) or as prescribed by law.

APPENDIX G
UKES
Guidelines for Good Practice
In Evaluation

Guidelines for Good Practice in Evaluation

Preface

You will find in these guidelines the embodiment of much discussion, debate and practice in the field of evaluation. The guidelines attempt to capture, in an easily assimilated way, a diverse range of principles and frameworks for action that have been useful for evaluations in a variety of contexts. Organised in different sections, the guidelines provide a reference point from different perspectives for a range of stakeholders involved in the evaluation process. In this way practice is encapsulated from the point of view of evaluators themselves, commissioners, participants (particularly those from whom views, ideas or experience are sought) and those engaged in self-evaluation in organisations. They are intended for use in evaluations taking place in any domain or discipline.

Writing guidelines to inform practice has real challenges in that it is easy to slide into a set of disembodied, somewhat rationalistic standards, which bear little resemblance to the ebb and flow of interactions as the social process of evaluation unfolds. Yet it is difficult to adopt a conversational tone that at the same time allows for quick reference and easy access. These guidelines are written in a matter-of-fact style that eschews jargon, the obscure and the insular. They are intended for use by the novice and the experienced alike. While inevitably they may be a little truncated, each statement is designed to offer a starting point for consideration by the stakeholders in an evaluation and to act as a reference for statements of evaluation ethics, intentions and generic practice. The guidelines are not definitive and will continuously evolve; they therefore should be received as work-in-progress.

The guidelines aim at neutrality in the sense that they provide frameworks for action that does not exemplify any particular evaluation approach. This aim is complex, ambitious and important. There is no evaluation stance for which these guidelines are inappropriate or inapplicable. Many of the statements have at their heart the need to be open and transparent about the expectations and requirements of all the stakeholders *whoever* they may be. As such the language used has striven to avoid hidden or tacit assumptions about the efficacy, dominance or normality of any single approach to evaluation.

The guidelines are prescriptive only in the sense that they rehearse what those engaged in the practical business of evaluation, from whatever perspective, have found to be both an honourable and effective way of interacting. We believe the guidance will come alive through use in the discussions and negotiations between people involved in evaluations. As such we hope the statements will promote conversation about evaluation in general but also support ways of negotiating some of the critical aspects of the evaluation process from commissioning to dissemination of evaluation findings

Introduction

These guidelines have been developed to support the work of evaluators, commissioners of evaluation, evaluation participants and those involved in self-evaluation. The guidelines are divided into four sections, each of which applies to one of four key interest groups. Whilst there is some crossover of good practice issues, they are addressed individually to ensure clarity and appropriateness to each one.

Guidelines for evaluators

Evaluators need to:

- *be explicit* about the purpose, methods, intended outputs and outcomes of the evaluation; be mindful of unanticipated effects and be responsive to shifts in purpose.
- *alert commissioners* to possible adjustments to the evaluation approach and practice; be open to dialogue throughout the process informing them of progress and developments.
- *consider whether* it is helpful to build into the contract forms of external support or arbitration (should the need arise).
- *have preliminary discussion/s with commissioners* prior to agreeing a contract.
- *adhere to the terms agreed in the contract* and consult with commissioners if there are significant changes required to the design or delivery of the evaluation.
- *demonstrate the quality of the evaluation* to other parties through progress reports e.g. on development and financial accountability and adhere to quality assurance procedures as agreed in the contract.
- *be aware of and make every attempt to minimise any potential harmful effects* of the evaluation prejudicing the status, position or careers of participants

Evaluators also need to:

- *demonstrate* that the evaluation design and conduct are transparent and fit for purpose.
- *demonstrate comprehensive and appropriate use of all the evidence* and that evaluation conclusions can be traced to this evidence.
- *work within the Data Protection Act* and have procedures which ensure the secure storage of data.
- *acknowledge intellectual property* and the work of others.
- *have contractual agreement over copyright* of evaluation methodology, findings, documents and publication.
- *write and communicate evaluation findings* in accessible language.
- *agree with commissioners from the outset about the nature of dissemination* in order to maximise the utility of the evaluation.

In practice evaluators need to:

- *demonstrate a commitment to the integrity of the process of evaluation and its purpose to increase learning in the public domain.*
- *be realistic about what is feasible to achieve and their capacity to deliver within the time-scale and budget agreed.*
- *know when to refuse or terminate an evaluation contract because it is undoable, self-serving, or threatens to undermine the integrity of the process.*
- *be prepared to argue the case for the public right to know in evaluation in specified contexts.*
- *treat all parties equally in the process of the evaluation and the dissemination of findings.*

Guidelines for commissioners

To ensure good practice in evaluation, it would be helpful if Commissioners:

- *acknowledge the benefits of external, independent evaluation.*
- *operate fair tendering situations in which competitors ideas are not exploited or intellectual property misused as a result of commissioning.*
- *hold preliminary consultations with all parties to the evaluation to support a relevant, realistic and viable specification.*
- *specify the purpose and audience(s) for the evaluation with appropriate background material to encourage relevant tenders.*
- *operate a tendering procedure that is open and fair ensuring that appropriately qualified assessors are involved, making explicit criteria upon which a tender decision will be made.*
- *clarify the constraints that commissioners operate under, e.g. timescales, budgets, deadlines, and accountability.*
- *adhere to the terms agreed in the contract and consult with evaluators and other interest groups if significant changes are required to the design or delivery of the evaluation.*
- *specify the legal terms and responsibilities of the evaluation in the contract.*
- *match the aims and potential outcome of the evaluation to the knowledge and expertise of the potential evaluator(s).*
- *provide access to documentation and data required for evaluation purposes.*
- *establish clear principles for the reporting and dissemination of evaluation reports funded by public*

monies, consistent with acknowledged procedures which ensure quality evaluation and reporting.

- *have realistic expectations on what an evaluation might provide* including sufficient time for evaluators to respond to an initial invitation to tender and produce a proposal.
- *include experienced evaluators (who are not potential applicants for funding) in initial drafts of evaluation specifications, including feasible budget and realistic timescales.*
- *have trust in evaluators and mutual respect between participants, commissioners and evaluator(s).*
- *take advice of evaluators on research methodologies for collecting and analysing data.*
- *communicate openly and have respect for people involved in the evaluation and keep the evaluation team informed of changes in circumstances affecting the evaluation.*
- *recognise where evaluators need to keep their sources of information anonymous.*
- *preserve the integrity of the findings, e.g. by not quoting or publicising such findings out of context.*

Guidelines for evaluation participants

All participants in an evaluation shall:

- *receive a proper explanation of the purpose and methods of the evaluation and should have opportunity to comment on how they are represented in the evaluation.*
- *receive an explanation of the evaluation agreement forming part of the negotiation of the evaluation teams access to a programme.*
- *have access to the evaluation team as agreed in the for purposes of feedback, reporting and ongoing support for the duration of the evaluation.*
- *have proper opportunity to be assured that the data they offer is consonant with the Data Protection Act and that any data made public is on the grounds of fairness, accuracy and relevance.*
- *be assured that in the event of a dispute or difficulties between evaluation participants and evaluators, they would have access to independent arbitration.*
- *be assured that evaluators have taken all reasonable measures to ensure that the reports are negotiated.* Final reports should normally be lodged in the public domain and made available to all participants. Reasons for exemptions need to be recorded.
- *have the right to be informed about the explicit use of interview transcripts or video of teaching events and are asked to offer their informed consent.*

Guidelines for self-evaluation

Participants in an institutional self-evaluation need to:

- *make the aims, objectives and purposes of the evaluation clear to all members of the institution.*

- *ensure* that the process is built into the structure and function of the institution.
- *have* a clear set of procedures for the sharing of data within and beyond the institution.
- *take steps to ensure* that all members of the institution believe the evaluation is worth doing.
- *acknowledge* that the sharing of knowledge and experience within the institution may be more threatening than to those outside and take steps to lessen this threat.
- *treat all colleagues equally* in the process of the evaluation and dissemination of findings.
- *ensure that all involved in the evaluation (whether as a data givers, collectors or users) are engaged* at some level from the start so they know what is happening and why.
- *adopt* methodologies that are economical and feasible to use in the time-scales and operations of the institution.

- *have the backing and support* of the head of the institution, including financial support, where appropriate for meetings, networking, dissemination and publication.
- *assure* members of the institution that the findings from the evaluation are fed back into development as well as providing a measure of accountability.
- *indicate* that the process is methodologically sound from which valid implications can be drawn for the precise purpose agreed.
- *ensure the agreement and understanding* of all members of the institution before starting the evaluation.
- *demonstrate consistency and predictability* of behaviour in the conduct and negotiation of the evaluation.
- *recognise and agree* when it is important to make data public and when, for the development of the institution, it is prudent to retain some data in confidence.
- *communicate openly and honestly* with colleagues, consistent with maintaining fair and equitable ethical procedures.
- *seek advice and /or consider* adopting a critical friend to conduct a process audit of the methodological rigor and fairness with which the evaluation is conducted.
- *communicate to colleagues in accessible language* and engage them in discussion on the utility of the evidence and findings.

APPENDIX H
Alcohol and Drug Costs

ALCOHOL AND DRUG COSTS

Based on:

ALCOHOL

PUB PRICES – Barmaid conversation 21/1/2001

Beer	£1.60 pt
Cider/lager	£1.80 pt
Bottles of above	£1.80
Spirits	£1.20 shot

OFFLICENCE – Based on Tesco prices 21/1/2001

Beer/lager	£0.90 500ml can
Cider	£2.00 2 litre bottle
Spirits	£9.00 standard bottle (70cl)
Sherry	£4.00 standard bottle (75cl)
Wine	£3.00 standard bottle (75cl)

ALCOHOL UNITS – Based on standard Health Education materials:

	Volume	Units	
BEER/LAGER	Ordinary Strength	½ pt 1	
		1 pt 2	
		1 can 1½	
	Export	1 pt 2½	
		1 can 2	
	Strong	½ pt 2	
		1 pt 4	
		1 can 3	
	Xtra strength	½ pt 2½	
		1 pt 5	
		1 can 4	
	CIDER	Average Strength	½ pt 1½
1 pt 3			
1 litre 5			
1 quart 6			
Strong		½ pt 2	
		1pt 4	
		1 litre 7	
		1 quart 8	
SPIRITS		On average	1 measure 1
			1 standard bottle 30
WINE	On average	1 glass 1	
		1 standard bottle 7	
		1 litre 10	
SHERRY	On average	1 measure 1	
		1 standard bottle 12	

OTHER DRUGS

DRUG SQUAD INFORMATION 24/1/2001

HEROIN

£110 1/2oz (3 1/2grams) – around £30 per gram

AMPHETAMINE

£10 per gram

£6 per tab

COCAINE

£35 per gram

ECSTASY

£6 per tab

LSD

£3.50 per tab

ILLICIT BENZOS

£0.50 per tab

CANNABIS RESIN

£70 per oz

Based on information drawn from evaluation respondents, Integra staff, and drugs agency information, and bearing in mind obvious caveats around:

**Resin quality and type,
Possible use of herbal instead,
Strength preferred of joints (3 or 5 skin), lungs, buckets, pipes etc,
Solo sessions or with friends,**

cost of cannabis use, unless otherwise estimated by respondents themselves (as at Time 2 interviews) is based on the following estimations:

1/4 oz yields anywhere between 8 “stiffies” and 16 “weak” joints. Taking the midpoint of 12 as an estimate, calculations here are therefore based 8x12 joints per oz, per £70:

1 joint is therefore estimated to cost around £0.75

1 oz yields anywhere between 20 and 70 buckets/lungs. Taking the midpoint of 45 as an estimate:

1 bucket/lung is estimated to cost around £1.50

APPENDIX I
Experiment, Quasi-experiment, or
Pre-experiment?

Throughout this thesis, the core of the INTEGRA evaluation design has been characterised as 'experimental'. To be more specific, it has been described as a 'pretest-posttest' experimental design, as insisted upon by my PhD external examiner. Whilst this is not a major issue for the thesis as it relates to the core design I inherited on appointment to the INTEGRA evaluation, and is therefore not central to my original contributions, I am, nevertheless, keen to be as lucid as reasonably possible.

As described in Rubin and Babbie (2001, p. 303), the essential components of experiments involve:

A/. Randomly assigning individuals to experimental and control groups and observation on dependent variables (e.g. self esteem of offenders);

B/. Introducing the independent variable (e.g. INTEGRA attendance for the experimental group and non-attendance for the control group);

C/. Comparing the amount of experimental and control group change on the dependent variables.

This is the classic pretest-posttest experimental design. In the case of the INTEGRA evaluation, however:

1/. There was no random assignment to experimental or control groups. Probationers were automatically referred to INTEGRA for assessment when their LSI-r (completed by their supervising Probation Officers) indicated they had problems associated with their alcohol and or other drug use and when they were also unemployed.

2/. There was no pre-INTEGRA observation on dependent variables. Respondents were interviewed usually after, or sometimes during, their INTEGRA participation at T1, and then followed-up 12 months later at T2. Experimental designs do exist which account for the lack of pre-test observation, or where researchers are concerned that a pre-test might have an impact on treatment effects. These are referred to as 'posttest-only' designs.

However, to be 'true' experiments they still require a suitable control group achieved by randomisation¹³.

3/. There was no equivalent control group. However, we were able to statistically equate users referred to the INTEGRA Project with those who were not referred (comparison group¹⁴), on measurable characteristics associated with some of our outcomes, by drawing 'controls' from the LSI-r database of Probationers. At my insistence, an attempt was made to deliver the evaluation interview to a random sample of these Probationers not referred to INTEGRA, but this was not successful. Three weeks of door-knocking resulted in only six completed interviews and increased doorstep antagonism towards myself.

The LSI-r analysis above was also useful in exploring possible differences between INTEGRA attenders who agreed to undertake the evaluation interview and those who, for whatever reason, did not. This was able to demonstrate contactability bias.

Perhaps for the INTEGRA evaluation core design to be 'truly' experimental it would have necessitated the random assignment of all probationers 'suitable' for INTEGRA (i.e. all those identified from their LSI-r as having alcohol or other drug problems and were unemployed) to experimental and control groups? However, the pragmatics of the Criminal Justice System, and the ethics of social research, would make this impractical. I understand that such a design was the subject of early discussions between the INTEGRA Evaluation Manager and the INTEGRA Programme Manager.

The term 'quasi-experiment' is used to describe impact/outcome designs (Campbell and Stanley, 1963; Cook and Campbell, 1979) where there are no randomly assigned control groups. The key issue here is how to control for threats to internal validity (and thus permit causal inferences) by constructing a suitable comparison group, or by statistically equating subjects to non-subjects on measurable characteristics associated with outcomes.

¹³ Randomisation is not the same as random sampling. Randomisation is a device for increasing internal validity; it does not seek to ensure that subjects are representative of a population, as in random sampling, but that experimental group subjects are representative of control group subjects.

¹⁴ The term 'comparison group' is commonly used instead of 'control group' when subjects are not assigned randomly.

On the other hand, the term 'pre-experiment' was coined by Campbell and Stanley (1963) to describe impact/outcome designs that do little to control for threats to internal validity. The simplest example of this design is the 'one-shot case study' (Rubin and Babbie, 2001). In this, a single group of subjects is measured on a dependent variable after the introduction of the intervention, and without comparing the results to anything else.

Another example is the 'posttest-only design with non-equivalent groups' (Cook and Campbell, 1979). This would entail assessing subjects on an outcome variable after they receive the intervention, and then comparing with another group who may plausibly be unlike the 'treated' subjects in any way.

All in all, perhaps it is safest to characterise the INTEGRA evaluation core design as having pretest-posttest experimental intent, and a 'mish-mash' of posttest-only pre-experimental and quasi-experimental application?

REFERENCES

- Abma, T. (1997). Sharing power, facing ambiguity. Evaluation and the postmodern dilemma. L. Mabry. London, JAI Press. 3: 105-119.
- Adelman, C. (1996). "Anything goes: Evaluation and relativism." Evaluation 2(3): 291-305.
- Altrichter, H. (1986). "Visiting two worlds: An excursion into the methodological jungle including an optional evening's entertainment at the Rigour Club." Cambridge Journal of Education 16(2): 131-143.
- Anderson, G. (1989). "Critical ethnography in education: origins, current status, and new directions." Review of Educational Research 59(3): 249-270.
- Andrews, D. A., Bonta, J.L. (1995). LSI-R Manual. North Tonawanda, NY, Multi-Health Systems.
- Andrews, D. (1995). The psychology of criminal conduct and effective treatment. What Works: Reducing reoffending. J. McGuire. Chichester, Wiley: 35-62.
- Anglin, M. D., Speckart, G. (1988). "Narcotics and Crime: a multi-sample, multi-method analysis." Criminology 26(2): 197-233.
- Apex Trust (1996). National Audit of Offender Employment Policies and Practice, Apex Trust.
- Barker, R. (1968). Ecological Psychology. Stanford, CA, Stanford University Press.
- Beauchamp, T. L., Childress, J.F. (2001). Principles of biomedical ethics (5th Edition). Oxford, Oxford University Press.
- Beecham, J. (1995). Collecting and estimating costs. The Economic Evaluation of Mental Health Care. M. Knapp. Aldershot, Arena.

Bennett, T. (1998). *Drugs and Crime: the results of research on drug testing and interviewing arrestees*. London, Home Office.

Bennett, T. (2000). *Drugs and Crime: the results of the second development stage of the NEW-ADAM programme*. London, Home Office.

Bennett, T., Holloway, K., Williams, T. (2001). *Drug use and offending: summary results of the first year of the NEW-ADAM research programme*. London, Home Office.

Bennington, G. (1996). Deconstruction. *The Blackwell Dictionary of Twentieth-Century Social Thought*. W. B. Outhwaite, T. Oxford, Blackwell.

Berger, P. L., Luckman, T. (1967). *The Social Construction of Reality*. New York, Free Press.

Best, D., Sidwell, C., Gossop, M., Harris, J., Strang, J. (2001). "Crime and Expenditure Among Polydrug Misusers Seeking Treatment." *British Journal of Criminology* 41: 119-126.

Bhaskar, R. (1997). *A Realist Theory of Science*. London, Verso.

Bickman, L., Ed. (1987). *Using Program Theory in Evaluation*. *New Directions for Program Evaluation Series (No. 33)*. San Francisco, Jossey-Bass.

Bloor, M. e. a. (1997). *Estimating the Prevalence of Injecting Drug Use and Serious Drug Use in Wales*. Cardiff, University College.

Bordin, E. S. (1980). *Of Human Bonds That Bind Or Free. Presidential address delivered at the meeting of the Society for Psychotherapy Research*. Pacific Grove, CA.

Bowman, M. L. (1989). "Testing Individual Differences in Ancient China." *American Psychologist* 44: 576-578.

Box, S. (1987). *Recession, Crime and Punishment*. London, Macmillan.

Brain, K., Parker, H., Bottomley, T. (1998). *Evolving Crack Cocaine Careers: New Users, Quitters and Long Term Combination Drug Users in N.W. England*. Manchester, University of Manchester.

Bridges, A. (1998). *Increasing The Employability Of Offenders: An Inquiry into Probation Service Effectiveness*. Oxford, University of Oxford Centre for Criminological Research in Association with Berkshire Probation Service.

Briggs, C. L. (1986). Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research. Cambridge, Cambridge University Press.

British Association of Social Workers (2002). *The Code of Ethics for Social Work*. Brimingham, BASW.

Bryman, A. (1988). Quantity and Quality in Social Research. London, Unwin Hyman.

Burlison, P., Schneider, J. (2002). *The INTEGRA Evaluation: Final Report*. Durham, Centre for Applied Social Studies, University of Durham.

Byrne, D. (1999). Social Exclusion. Buckingham, Open University Press.

Cabinet Office (1999). *Professional Policy Making for the Twenty First Century*. London, Strategic Policy Team: Cabinet Office.

Cabinet Office (2000). *Adding it up: Improving Analysis and Modelling in Central Government*. London, Central Office of Information.

Cacciola, J. S., Koppenhaver, J.M., McKay, J.R., Altman, A.I. (1999). "Test-retest reliability of the lifetime items on the addiction severity index." Psychological Assessment 11: 86-93.

Cahoone, E., Ed. (1996). From Modernism to Postmodernism: an anthology. Oxford, Blackwell.

Campbell, D. T. (1957). "Factors relevant to the validity of experiments in social settings." Psychological Bulletin(54): 297-312.

Campbell, D. T. (1960). "Blind variations and selective retention in creative thought as in other knowledge processes." Psychological Review(67): 380-400.

Campbell, D. T., Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago, Rand McNally.

Campbell, D. T. (1974). Evolutionary epistemology. The philosophy of Karl Popper. P. A. Schilpp. La Salle, IL, Open Court: 413-463.

Campbell, D. T. (1979). Degrees of freedom and the case study. Qualitative and Quantitative Methods in Evaluation Research. T. a. R. Cook, C. Beverly Hills, CA, Sage.

Campbell, D. T. (1984a). Hospital and landsting as continuously monitoring social polygrams: Advocacy and warning. Evaluation of mental health service programs. B. a. v. K. Cronholm, L. Stockholm, Forskningsraadet Medicinska: 13-39.

Campbell, D. T. (1984b). Can we be scientific in applied social science? Evaluation Studies Review Annual. R. F. Connor, Altman, D.G., and Jackson, C. Beverly Hills, CA, Sage. 9: 26-48.

Campbell, D. T. (1988). Methods for the Experimenting Society. Methodology and Epistemology for Social Science: Selected Papers. E. S. Overman. Chicago, University of Chicago Press.

Campbell, D. T. (1994). Foreward. Case Study Research. R. K. Yin. Thousand Oaks, Sage.

Caracelli, V. J., Greene, J.C. (1997). Crafting Mixed-Method Evaluation Designs. New Directions for Program Evaluation. J. C. Greene, Caracelli, V.J. San Francisco, Jossey-Bass. 74: 19-32.

Carey, J. W., Wenzel, P.H., Reilly, C., Sheridan, J., Steinberg, J.M. (1998). "CDC EZText: Software for Management and Analysis of Semistructured Qualitative Data Sets." Cultural Anthropology Methods 10(1): 14-20.

Chapman, T., Hough, M. (1998). Evidence Based Practice: A Guide to Effective Practice. London, Home Office: HM Inspectorate of Probation.

Chelimsky, E. (1995). "The Political Environment of Evaluation and What It Means for the Development of the Field." Evaluation Practice 16(3): 215-225.

Chelimsky, E. (1997). The Coming Transformations in Evaluation. Evaluation for the 21st Century. E. a. S. Chelimsky, W.R. Thousand Oaks, CA, Sage.

Chelimsky, E., Shadish, W.R., Ed. (1997). Evaluation for the 21st Century: A Handbook. Thousand Oaks, CA, Sage.

Chen, H.-T. (1990). Theory-Driven Evaluations. Thousand Oaks, CA, Sage Publications.

Clarke, S. (1999). "Justifying Deception in Social Science Research." Journal of Applied Philosophy 16(2): 151-166.

Coid, J., Carvell, A., Kittler, Z., Healey, A., Henderson, J. (2000). Opiates, Criminal Behaviour, and Methadone Treatment. London, Home Office.

Comstock, D. (1982). A method for critical research. Knowledge and Values in Educational Research. F. a. F. Bredo, W. Philadelphia, Temple University Press.

Cook, T. D., Campbell, D.T. (1979). Quasi-Experimentation Design and Analysis Issues for Field Settings. Skokie, IL, Rand McNally.

Cook, T. D. (1997). Lessons learned in evaluation over the past 256 years. Evaluation for the 21st Century: a Handbook. E. Chelimsky, Shadish, W.R. London, Sage: 30-52.

Corrigan, P. W., Faber, D., Rashid, F., Leary, M. (1999). "The construct validity of empowerment among consumers of mental health services." Schizophrenia Research 38(1): 77-84.

Cronbach, L. J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., Weiner, S.S. (1980). Toward Reform of Program Evaluation. San Francisco, Jossey-Bass.

Davies, P. (1997). Women, Crime and an Informal Economy: Female Offending and Crime for Gain. British Criminology Conference, Queens University, Belfast.

de Jong, M., Schellens, P.J. (1997). "Reader-Focused Text Evaluation: an Overview of Goals and Methods." Journal of Business and Technical Communication 11(4): 402-432.

de Jong, M., Schellens, P.J. (1998). "Focus Groups or Individual Interviews? A Comparison of Text Evaluation Approaches." Technical Communication(First Quarter): 77-88.

De Li Periu, H., MacKenzie, D. (2000). "Drug involvement, lifestyles and criminal activities among probationers." Journal of Drug Issues 30(3): 593-620.

Denzin, N. (1989). The Research Act. New York, McGraw Hill.

Dhiri, S., Brand, S. (1999). Crime Reduction Programme Guidance Note 1. Analysis of costs and benefits: guidance for evaluators. London, Home Office.

Dilthey, W. (1894 (1977)). Descriptive Psychology and Historical Understanding. The Hague, Martinus Nijhoff.

Dobinson, I., Ward, P. (1986). "Heroin and Property Crime: An Australian Perspective." Journal of Drug Issues 16(2): 249-262.

Downes, D. (1993). Employment Opportunities For Offenders. London, Home Office.

Drane, J. F. (1982). Ethics and psychotherapy: A philosophical perspective. Ethics and values in psychotherapy. M. Rosenbaum. New York, Free Press: 15-50.

Edmunds, M., May, T., Hough, M., Hearnden, I. (1998). Arrest Referral: Emerging Lessons from Research. London, Home Office.

Edmunds, M., Hough, M., Turnbull, P.J., May, T. (1999). Doing Justice to Treatment: referring offenders to drug services. London, Home Office.

Eng, G., Israel, B (1994). "Empowerment." Health Education Quarterly 21(2 and 3).

Etzioni, A. (1996). The New Golden Rule: Community and Morality in a Democratic Society. New York, Basic Books.

Etzioni, A. (1999). The Limits of Privacy. New York, Basic Books.

Everitt, A. (1996). "Developing critical evaluation." Evaluation(2): 173-188.

Farrall, S. (1995). "Why Do People Stop Offending?" The Scottish Journal Of Criminal Justice Studies 1(1): 51-59.

Farrington, D., Gallagher, B., Morley, L., St. Ledger, R., West, D. (1986). "Unemployment, School Leaving and Crime." British Journal of Criminology 26(4).

Fetterman, D. (1994). "Empowerment evaluation." Evaluation Practice(15): 1-16.

Fink, A. (1998). Conducting Research Literature Reviews: From paper to the Internet. Thousand Oaks: CA, Sage Publications

Fitzgerald, J., Hamilton, M. (1996). "The consequences of knowing: ethical and legal liabilities in illicit drug research." Sociology of Science and Medicine 43(11): 1591-1600.

Flood-Page, C., Campbell, S., Harrington, V., Miller, J. (2000). Youth Crime: findings from the 1998/99 Youth Lifestyles Survey. London, Home Office.

Fowler, F. J., Mangione, T.W. (1990). Standardized Survey Interviewing. Newbury Park, CA, Sage.

Fry, C., Dwyer, R. (2001). "For love or money? An exploratory study of why injecting drug users participate in research." Addiction 96(9): 1319-1325.

Fuller, R., Petch, A. (1995). Practitioner Research: The Reflexive Social Worker. Buckingham, Open University Press.

Gendreau, P., Little, T., Goggin, C. (1995). A Meta-Analysis of the Predictors of Adult Offender Recidivism: Assessment Guidelines for Classification and Treatment. Saint John, NB, University of New Brunswick.

Gilligan, C. (1982). In a different voice. Cambridge, MA, Harvard University Press.

Glaser, B. G., Strauss, A.L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, Aldine.

Goldstein, P. (1985). "The drugs/violence nexus: A tripartite conceptual framework." Journal of Drug Issues 15(4): 493-506.

Gossop, M., Marsden, J., Stewart, D. (1998). NTORS at One Year. The National Treatment Outcome Research Study. Changes in Substance Use, Health and Criminal Behaviour at One Year after Intake. London, Department of Health.

Graham, J., Bowling, B. (1996). Young People and Crime. London, Home Office Research Study.

Greene, J. C. (1990). Three views on the nature and role of knowledge in social science. The Paradigm Dialog. E. Guba. Newbury Park, Sage.

Greene, J. C. (1992). "The practitioner's perspective." Curriculum Inquiry 21(1): 39-45.

Greene, J. C. (1995). Evaluators as advocates. Annual Meeting of the American Evaluation Association, Vancouver.

Greene, J. C. (1997). Participatory Evaluation. Evaluation and the Postmodern Dilemma. L. Mabry. London, JAI Press. 3: 171-189.

Griffiths, R. (1973). "A Standardised assessment of work behaviour of psychiatric patients." Journal of Psychiatry(123): 403-408.

Guba, E. G., Lincoln, Y.S. (1989). Fourth Generation Evaluation. Newbury park: CA, Sage Publications.

Guba, E. (1990). The Alternative Paradigm Dialog. The Paradigm Dialog. E. Guba. Newbury Park, Sage.

Hammersley, R. H., Forsyth, A., Morrison, V., Davies, J.B. (1989). "The Relationship Between Crime and Opioid Use." British Journal of Addiction 84(9): 1029-1043.

Hammersley, M. (1996). The relationship between qualitative and quantitative research: Paradigm loyalty versus methodological eclecticism. Handbook of Qualitative Research Methods for Psychology and the Social Sciences. J. T. E. Richardson. Leicester, The British Psychological Society.

Harre, R. (1972). The philosophies of science. Oxford, Oxford University Press.

Harris, S., Mowbray, C., McCrohan, N., Greenfield, A. (1993). Work Behaviors and Attitudes Scale (WBAS). Michigan.

Harrison, L. D. (1992). "The drug-crime nexus in the USA." Contemporary Drug Problems 19(2): 181-202.

Hart, E., Bond, M. (1995). Action Research for Health and Social Care: a guide to practice. Buckingham, Open University Press.

Hearnden, I., Harocopos, A. (2000). Problem Drug Use and Probation in London. London, Home Office.

Hodgson, R. (1994). "Treatment of Alcohol Problems." Addiction 89: 1529-1534.

Holder, H., Longhaugh, R., Miller, W.R., Rubonis, A.V. (1991). "The cost effectiveness of treatment for alcohol problems: a first approximation." Journal of Alcohol Studies 52: 517-540.

Homan, R. (1991). The Ethics of Social Research. London, Longman.

Home Office (1999). Reducing crime and tackling its causes: A briefing note on the Crime Reduction Programme. London, Home Office.

Home Office (2000). National Standards for the Supervision of Offenders in the Community. London, Home Office.

Horvath, A. O., Greenberg, L.S. (1989). "Development and Validation of the Working Alliance Inventory." Journal of Counseling Psychology 36(2): 223-233.

Hough, M. (1996). Drugs misuse and the criminal justice system: a review of the literature. London, Home Office.

House, E. R. (1980). Evaluating with Validity. Beverly Hills, CA, Sage.

House, E. R. (1991). Evaluation and Social Justice: Where are we? Evaluation and education: At quarter century. M. M. a. D. C. Philips. Chicago, University of Chicago Press: 233-247.

House, E. R. (1993). Professional Evaluation. Newbury Park, CA, Sage Publications.

House, E. R., Howe, K.R. (1999). Values in Evaluation and Social Research. Thousand Oaks, CA, Sage Publications.

House of Commons Select Committee on Home Affairs, S. C. o. H. (1998). Alternatives to prison: Third Report - List of Conclusions and Recommendations. London, House of Commons

Hser, Y. I., Maglone, M., Polinsky, M., Angun, M.D. (1998). "Predicting drug treatment entry among treatment seeking individuals." Journal of Substance Abuse Treatment 15(3): 213-20.

Hughes, R. (1999). "Why do people agree to participate in social research? The case of drug injectors." International Journal of Social Research Methodology 1(4): 315-324.

Humphries, B. (1999). Feminist Evaluation. Evaluation and Social Work Practice. I. a. L. Shaw, J. London, Sage.

Jarvis, G., Parker, H. (1989). "Young Heroin Users and Crime: How do the "New Users" finance their habits?" British Journal of Criminology 29(2): 175-185.

Kant, I. (1956). Critique of practical reason. New York, Liberal Arts Press.

Kazi, M. A. F. (2000). Realist Evaluation for Evidence-Based Practice. Fourth Conference of the European Evaluation Society, Lausanne, Switzerland,.

Kazi, M. A. F. (2000). "Contemporary Perspectives in the Evaluation of Practice." British Journal of Social Work (30): 755-768.

Kazi, M.A.F., Smith, M., Clarke, B., Manby, M. (2002). "Realist Evaluation of Young People with Sexually Abusive Behaviour", Centre for Evaluation Studies, Huddersfield, England: University of Huddersfield.

Kitchener, K. S. (1984). "Intuition, critical evaluation and ethical principles: The foundation for ethical decisions in counseling psychology." The Counseling Psychologist 12(3): 43-56.

Kivlighan, D. M. (1990). "Relation between counselors' use of intentions and clients' perception of working alliance." Journal of Counseling Psychology(37): 27-32.

Kokotovic, A. M., Tracey, T.J. (1990). "Working alliance in the early phase of counseling." Journal of Counseling Psychology(37): 16-21.

Krosnick, J. A. (1999). "Survey Research." Annual Review of Psychology(50): 537-567.

Kushner, S. (2000). Personalizing Evaluation. London, Sage.

Lader, D., Singleton, N., Meltzer, H. (2000). Psychiatric Morbidity Among Young Offenders in England and Wales. London, ONS.

Lather, P. (1991). Getting Smart: Feminist Research and Pedagogy with/in the Post-modern. New York, Routledge.

Latour, B. (1988). The politics of explanation: an alternative. Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge. S. Woolgar. London, Sage.

Lee, R. M. (1995). Doing Research on Sensitive Topics. London, Sage.

Levitas, R. (1998). The Inclusive Society? Basingstoke, Macmillan.

Lincoln, Y. S., Guba, E.G. (1985). Naturalistic Inquiry. Beverly Hills, CA, Sage.

Lincoln, Y., Guba, E. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. Naturalistic Evaluation. D. D. Williams. San Francisco, Jossey-Bass. 30.

Lincoln, Y., Guba, E. (1986a). "Research, evaluation and policy analysis: heuristics and disciplined inquiry." Policy Studies Review 5(3): 546-565.

Lincoln, Y. (1990). The making of a constructivist: a remembrance of transformations past. The Paradigm Dialog. E. Guba. Newbury Park, Sage.

Lipsey, M. (1995). What Do We Learn From 400 Research Studies On The Effectiveness Of Treatment With Juvenile Delinquents? What Works In Reducing Offending: Guidelines from Research and Practice. J. McGuire. Chichester, John Wiley and Sons.

Lipsky, M. (1980). Street-level Bureaucracy. Dilemmas of the Individual in Public Services. New York, Russell Sage Foundation.

Lister, R., Ed. (1996). Charles Murray and the Underclass. London, IEA.

Lister, R. (2002). "A Politics of Recognition and Respect: Involving People with Experience of Poverty in Decision Making that Affects their Lives." Social Policy and Society 1(1): 37-46.

Lloyd, C., Mair, G., Hough, M. (1994). Explaining Reconviction Rates: A Critical Analysis. Home Office Research Study No. 136. London, HMSO.

Lord, G. (1986). The Broadwater Farm Inquiry: Report of the Independent Inquiry into Disturbances of October 1985 at the Broadwater Farm Estate, Tottenham. London, Karla Press.

Mabry, L., Ed. (1997). Evaluation and the Postmodern Dilemma. Advances in Program Evaluation. London, JAI Press.

Mabry, L. (1997a). Implicit and Explicit Advocacy in Postmodern Evaluation. Evaluation and the Postmodern Dilemma. L. Mabry. London, JAI Press. 3: 191-203.

Mabry, L. (1997b). A Postmodern Text on Postmodernism? Evaluation and the Postmodern Dilemma. L. Mabry. London, JAI Press. 3: 1-19.

Macdonald, G., Veen, C., Tones, K. (1996). "Evidence for success in health promotion: suggestions for improvement." Health Education Research, Theory and Practice 11(3): 367-376.

MacGregor, S. (2000). "The Drugs-Crime Nexus." Drugs: education, prevention and policy 7(4): 311-316.

MacLure, M. (1999). Discursive (il)literacy in research, policy and practice. Annual meeting of the British Educational Research Association, Brighton, UK.

Madaus, G. F., Stuffebeam, D.L., Scriven, M.S. (1983). Program Evaluation: A historical overview. Evaluation Models: Viewpoints on educational and human services evaluation. G. F. Madaus, and D.L. Stuffebeam. Boston, Kluwer-Nijhoff: 3-22.

Maden, A., Swinton, M., Gunn, J. (1991). "Drug Dependence in Prisons." British Medical Journal 302(6781): 880.

Mair, G., Nee, C. (1992). "Day Centre Reconviction Rates." British Journal of Criminology 32: 329-39.

Marlatt, G. A., Gordon, J.R. (1985). Relapse prevention. London, Guilford Press.

Marlowe, D. B., Kirby, K.C., Bonieshire, L.M., Glass, D.J., Dodds, L.D., Husband, S.D., Platt, J.J., Festinger, D.S. (1996). "Assessment of coercive and non-coercive pressures to attend drug abuse treatment." Drug and Alcohol Dependence(42): 77-84.

Marrow, A. J. (1969). The practical theorist: The life and work of Kurt Lewin. New York, Basic Books.

Marsden, J. e. a. (1998). "The Maudsley Addiction Profile (MAP): A Brief Instrument for Assessing Treatment Outcome." Addiction 93(12): 1857-1867.

Maykut, P., Morehouse, R. (1994). Beginning Qualitative Research: A philosophic and practical guide. Bristol, PA, The Falmer Press.

Mays, N., Pope, C., Ed. (1996). Qualitative Research in Health Care. London, BMJ Publishing.

McGuire, J., Ed. (1995). What Works: Reducing Offending. Chichester, Wiley.

Mclvor, G. (1995). Practitioner evaluation in probation. What Works: Reducing reoffending. J. McGuire. Chichester, Wiley: 209-220.

Measham, F., Aldridge, J., Parker, H. (2001). Dancing on Drugs. Risk, Health and Hedonism in the British Club Scene. London, Free Association Books.

Mendelow, A. L. (1987). Stakeholder Analysis for Strategic Planning and Implementation. Strategic Planning and Management Handbook. W. R. C. King, D.I. New York. Van Nostrand Reinhold: 176-191.

Merrington, S., Hine, J. (2001). Handbook for Evaluating Probation Work with Offenders. London, Home Office, HMIP, ACOP, NPRIE.

Mishler, E. G. (1986). Research Interviewing. Cambridge, MA, Harvard University Press.

Mowbray, C. T. (1995). "Predictors of Work Status and Future Work Orientation in People with a Psychiatric Disability." Psychiatric Rehabilitation(Fall): 17-28.

Nellis, M. (1995). "Probation Partnerships, voluntary action and community justice." Social Policy and Administration 29(2): 91-109.

Netten, A., Beecham, J., Ed. (1993). Costing Community Care: Theory and Practice. Aldershot, Arena.

Newman, D. L., Brown, R.D. (1996). Applied Ethics for Program Evaluation. Thousand Oaks, Sage Publications Inc.

Nielson, J. M. (1990). Introduction. Feminist Research Methods. J. Nielson. Boulder, Co, Westview Press.

Noddings, N. (1984). Caring: A feminine approach to ethics and moral education. Berkeley, University of California Press.

Nutley, S., Davies, H. (2000). Criminal Justice: using evidence to reduce crime. What Works? Evidence-based policy and practice in public services. S. N. H. Davies, and P. Smith. Bristol. The Policy Press: 93-116.

Oldfield, M. (1996). The Kent Reconviction Study. Maidstone, Kent Probation Service.

Outhwaite, W. (1987). New Philosophies of Social Science: Realism, Hermeneutics and Critical Theory. Basingstoke, Macmillan.

Parker, H., Kirby, P. (1996). Methadone Maintenance and Crime Reduction on Merseyside. Police Research Group, Crime Prevention Series. London, Home Office.

Parker, H., Aldridge, J., Measham, F. (1998). Illegal Leisure: The Normalisation of Adolescent Recreational Drug Use. London, Routledge.

Parlett, M., Hamilton, D. (1977). Evaluation as Illumination: A new approach to the study of innovatory programs. Beyond The Numbers Game. D. Hamilton, Jenkins, D., King, C., MacDonald, B., Parlett, M. Basingstoke, Macmillan.

Pattison, S. (1999). "Are Professional Codes Ethical?" Counselling(December): 374-380.

Patton, M. Q. (1978). Utilization-Focused Evaluation. Beverly Hills, CA, Sage.

- Patton, M. Q. (1982). Practical Evaluation. Beverly Hills, CA, Sage.
- Patton, M. Q. (1986). Utilization-Focused Evaluation. Beverly Hills, CA, Sage.
- Patton, M. Q. (1988). "The evaluator's responsibility for utilization." Evaluation Practice 9(1): 5-24.
- Patton, M. Q. (1997). Utilization-Focused Evaluation. Thousand Oaks, CA, Sage Publications Inc.
- Pawson, R., Tilley N. (1997). Realistic Evaluation. Thousand Oaks, CA, Sage Publications Inc.
- Pettigrew, M. (2002). "Some Reflections on Evaluation in the European Commission." The Evaluator(11): 12.
- Poole, R. (1972). Towards Deep Subjectivity. London, Allen Lane, Penguin.
- Popkewitz, T. (1990). Whose Future? Whose past? The Paradigm Dialog. E. Guba. Newbury Park, Sage.
- Power, R. (1989). "Participant observation and its place in the study of illicit drug abuse." British Journal of Addiction 84(1): 43-52.
- Prochaska, J. O., DiClemente, C.C. (1984). The Transtheoretical Approach: crossing traditional boundaries of therapy. Homewood, Ill, Dow Jones/Irwin.
- Ramsey, M., Baker, P., Goulden, C., Sharp, C., Sondhi, A. (2001). Drug Misuse Declared in 2000: results from the British Crime Survey. London, Home Office.
- Rawls, J. (1971). A theory of justice. Cambridge, MA., Harvard University Press.
- Raynor, P. (1998). "Attitudes, Social Problems and Reconvictions in the 'STOP' Probation Experiment." The Howard Journal 37(1): 1-15.
- Raynor, P., Kynch, J., Roberts, C., Merrington, S. (2000). Risk and Need Assessment in Probation: an evaluation. London, Home Office.

Reason, P., Rowan, J., Ed. (1981). Human inquiry. A sourcebook of new paradigm research. Chichester, John Wiley.

Reiben, C. (1996). "Participatory evaluation of development assistance: Dealing with power and facilitative learning." Evaluation(2): 151-172.

Reinharz, S. (1983). Experiential analysis: a contribution to feminist research. Theories of Women's Studies. G. D. K. Bowles, R. London, Routledge and Kegan Paul.

Reinharz, S. (1993). Experimental Analysis: A Contribution to Feminist Research. Theories of Women's Studies. G. B. A. R. Klien. Boston, Routledge and Kegan Paul.

Rice, G., Thomas, T. (2000). The Possibilities for a Confidential Drug Service. Human Rights and Penal Issues: 18th Annual Conference of the Howard League for Penal Reform, New College, Oxford.

Richards, M. (1996). Evaluation of the Intensive Probation Centre. Cambridge, Cambridgeshire Probation Service.

Robinson, I. (1991). "Confidentiality for whom." Soc. Sci. Med. 32(279).

Robson, C. (1993). Real World Research. Oxford, Blackwell.

Robson, C. (2000). Small-Scale Evaluation. London, Sage.

Rogers, E. S., Chamberlin, J., Ellison, M.L., Crean, T. (1997). "A Consumer Constructed Empowerment Scale to Measure Empowerment Among Users of Mental Health Services." Psychiatric Services 48(8): 1042-1047.

Rorty, R. (1991). Objectivity, Relativism, and Truth. Cambridge, Cambridge University Press.

Rosenblum, A. (1995). "Treatment intensity and reduction in drug use for cocaine dependent methadone patients." Journal of Psychoactive Drugs 27(2): 151-159.

Roshier, B. (1995). A Comparative Study of Reconviction Rates in Cleveland. Durham, University of Durham.

Rossi, P. H., Freeman, H.E., Lipsey, M.W. (1999). Evaluation: A Systematic Approach. Thousand Oaks, CA, Sage Publications.

Rumgay, J. C., S. (1998). "Pitfalls and Prospects in Partnership: Probation Programmes for Substance Misusing Offenders." The Howard Journal 37(2): 124-136.

Rubin, A., Babbie, E. (2001). Research Methods for Social Work. Belmont, CA: Wadsworth.

Salasin, S. (1974). "Exploring goal-free evaluation: An interview with Michael Scriven." Evaluation(2): 9-16.

Sarno, C., Hearnden, I., Hedderman, C., Hough, M., Nee, C., Herrington, V. (2000). Working their way out of offending: an evaluation of two probation employment schemes. London, Home Office.

Schober, M. F., Conrad, F.G. (1997). "Does conversational interviewing reduce survey measurement error?" Public Opinion Quarterly(61): 576-602.

Scriven, M. S. (1976). Evaluation bias and its control. Evaluation Studies review annual. G. V. Glass. Beverly Hills, CA, Sage. 1: 101-118.

Scriven, M. S. (1980). The Logic of Evaluation. Inverness, CA, Edgepress.

Scriven, M. S. (1983). Evaluation Ideologies. Evaluation models: Viewpoints on educational and human services evaluation. G. F. Madaus, Scriven, M., and Stufflebeam, D.L. Boston, Kluwer-Nijhoff: 229-260.

Scriven, M. S. (1986). "New frontiers of evaluation." Evaluation Practice 7(1): 7-44.

Scriven, M. S. (1991). Evaluation Thesaurus. Newbury Park, CA, Sage Publications.

Scriven, M. (1991). Beyond Formative and Summative Evaluation: At Quarter Century. 90th Yearbook of the National Society for the Study of Education. M. W. A. P. McLaughlin, D.C. Chicago, University of Chicago Press.

Scriven, M. S. (1995). The logic of evaluation and evaluation practice. Reasoning in Evaluation: Inferential Links and Leaps. D. Fournier. San Francisco, American Evaluation Association/Jossey-Bass.

Scriven, M. S. (1997). Truth and objectivity in evaluation. Evaluation for the 21st Century. E. a. S. Chelimsky, W.R. Thousand Oaks, Sage: 477-500.

Sechrest, L. (2000). Unobtrusive Measures, Sage Publications.

Shadish, W. R., Epstein, R. (1987). "Patterns of Program Evaluation Practice Among Members of the Evaluation Research Society and Evaluation Network." Evaluation Review 11: 555-90.

Shadish, W. R. J., Cook, T.D., Leviton, L.C. (1991). Foundations of Program Evaluation. Newbury Park, CA, Sage.

Shadish, W. R., Newman, D.L., Scheirer, M.A., Wye, C. (1995). Guiding Principles for Evaluators. New Directions for Program Evaluation (66). San Francisco, Jossey-Bass.

Shapland, J. (1997). Looking at Opportunities in the Informal Economy of Cities. European Seminar on the Informal Economy, Freiburg i.Br.

Shaw, I. F. (1999). Qualitative Evaluation. London, Sage.

Sieber, J. E. (1998). Planning ethically responsible research. Handbook of Applied Social Research Methods. D. J. R. L. Bickman. Thousand Oaks, CA, Sage.

Singleton, N., Farrel, M., Meltzer, H. (1999). Substance Misuse Among Prisoners in England and Wales. London, ONS.

Smith, N. L., Ed. (1981). Metaphors for Evaluation: Sources of New Methods. Beverly Hills, CA, Sage.

Sohm, B. (1994). "The deconstruction of actuality: An interview with Jacques Derrida." Radical Philosophy(Autumn): 28-41.

Stake, R. E. (1975). To evaluate an arts program. Evaluating the arts in education: A responsive approach. R. E. Stake. Columbus, OH, Merrill: 13-31.

Stake, R. E. (1978). "The case study method in social inquiry." Educational Researcher(7): 5-8.

Stake, R. E., Trumbull, D.J. (1982). "Naturilistic generalizations." Review Journal of Philosophy and Social Science.(7): 1-12.

Stake, R. E. (1991). Retrospective on 'The Countenance of Educational Evaluation'. Evaluation and Education at Quarter Century. M. a. P. McLaughlin, D. Chicago, University of Chicago Press.

Stake, R. E. (1997). The Fleeting Discernment of Quality. Evaluation and the Postmodern Dilemma. L. Mabry. London, JAI Press. 3: 41-59.

Stern, E. (1999). "Why Parliament Should Take Evaluation Seriously." Parliamentary Review.

Stronach, I. (1997). Evaluation With The Lights Out: Deconstructing illuminative evaluation and new paradigm research. Evaluation and the Postmodern Dilemma. L. Mabry. London, JAI Press. 3: 21-39.

Stronach, I., MacLure, M. (1997). Educational Research Undone: The Postmodern Embrace. Buckingham, Open University Press.

Suchman, E. A. (1967). Evaluative research: Principles and practice in public service and social action programs. New York, Russell Sage Foundation.

Suchman, L., Jordan, B. (1990). "Interactional troubles in face-to-face survey interviews." Journal of American Statistical Association(85): 232-53.

Suchman, L., Jordan, B. (1992). Validity and the collaborative construction of meaning in face-to-face surveys. Questions About Questions. J. Tanur. New York, Sage.

Summers, Z. (2002). Coercion and the Criminal Justice System. Working With Substance Misusers: A Guide to Theory and Practice. T. Petersen, McBride, A. London, Routledge.

Swigonski, M. (1993). "Feminist standpoint theory and questions of social work research." Affilia 8(2): 171-183.

Tarling, R. (1982). "Unemployment and Crime." Research Bulletin(14): 28-33.

Turnbull, P. (1999). Drug Treatment and Testing Orders - interim evaluation. London, Home Office.

Turnbull, P. J., McSweeney, T., Hough, M., Webster, R., Edmunds, M. (2000). Drug Treatment and Testing Orders: Final Evaluation Report. London, Home Office.

UKES (2002). "UK Evaluation Society: Guidelines for Good Practice in Evaluation." The Evaluator(11): 5-8.

VanderPlaat, M. (1995). "Beyond technique: Issues in evaluating for empowerment." Evaluation(1): 81-96.

Wadsworth, Y. (1991). Everyday Evaluation on the Run. St. Leonards, Allen and Unwin.

Wadsworth, Y. (1997). Everyday Evaluation on the Run. St. Leonards, Allen and Unwin.

Wadsworth, Y. (1997). Do It Yourself Social Research. St. Leonards, Allen and Unwin.

Wakefield, N. (1990). Postmodernism: The Twilight of the real. London, Pluto Press.

Walters, G. (1998). Changing Lives of Drugs and Crime. Chichester, Wiley and Sons.

Weisbrod, B., Stein, M., Test, L. (1980). "Alternatives to mental hospital treatment: Economic Benefit-cost analysis." Archives of General Psychiatry(37): 400-405.

Weiss, C. H. (1972). Evaluation research: Methods for assessing program effectiveness. Englewood Cliffs, NJ, Prentice-Hall.

Weiss, C. H. (1973). "Where politics and evaluation research meet." Evaluation(1): 37-45.

Weiss, C. H. (1977). "Research for policy's sake: The enlightenment function of social research." Policy Analysis(3): 531-545.

Weiss, C. H. (1980). "Knowledge creep and decision accretion." Knowledge, Creation, Diffusion, Utilisation 1(3): 381-404.

Weiss, C. H., Bucuvalas, M.J. (1980). Social science research and decision-making. New York, Columbia University Press.

Weiss, C. H. (1982). "Policy research in the context of diffuse decision making." Journal of Higher Education(53): 619-639.

Weiss, C. H. (1983). Ideology, interest, and information: The basis of policy decisions. Ethics, the social sciences, and policy analysis. D. Callahan, Jennings, B. New York, Plenum: 213-245.

Weiss, C. H. (1987a). "The circuitry of enlightenment." Knowledge: Creation, Diffusion, Utilisation 8: 274-81.

Weiss, C. H. (1987b). "Evaluating social programs: What have we learned?" Society 25(1): 40-45.

Weiss, C. H. (1988). "If programme decisions hinged only on information." Evaluation Practice 9(1): 15-28.

Weiss, C. H. (1995). Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts. J. E. A. Connell. Washington DC, The Aspen Institute.

Weiss, C. H. (1998). Evaluation: methods for studying programs and policies (2nd Ed). Upper Saddle River, NJ., Prentice Hall.

Whitmore, E. (1994). To tell the truth. Working with oppressed groups in participatory approaches to inquiry. Participation in human inquiry. P. Reason. Newbury Park, CA, Sage.

Wholey, J. S. (1985). "Managing for high performance: The role of evaluation." Evaluation News(6): 40-50.

Wholey, J. S., Ed. (1987). Organizational Excellence: Stimulating quality and communicating value. Lexington, MA., Lexington Books.

Wilkinson, J., Morgan, D. (1995). The Impact of Ilderton Motor Project on Motor Vehicle Crime and Offending. London, Inner London Probation Service.

Winckler, J. (1987). The fly on the wall of the inner sanctum: observing company directors at work. Research Methods for Elite Studies. G. a. W. Moysen, M. London, Allen & Unwin.

W.K. Kellogg Foundation (1998). Evaluation Handbook. Battle Creek, MI, W.K. Kellogg Foundation.

W.K. Kellogg Foundation, W. K. K. (2001). Logic Model Development Guide. Battle Creek, MI., W.K. Kellogg Foundation.

Wolf, M. (1992). A thrice-told tale: Feminism, postmodernism, and ethnographic responsibility. Stanford, CA., Stanford University Press.

Wowra, S. A., McCarter, R. (1999). "Validation of the Empowerment Scale With an Outpatient Mental Health Population." Psychiatric Services 50(7): 959-961.

Wright, L. (1999). Young People and Alcohol. A Literature Review. London, Health Education Authority.

Zimmerman, M. A. (1995). "Psychological empowerment: Issues and illustrations." American Journal of Community Psychology 23(5): 581-599.

