



Realist Evaluation in Practice

Health and Social Work

MANSOOR A.F. KAZI

Realist Evaluation in Practice

Health and
Social Work

MANSOOR A.F. KAZI



SAGE Publications
London • Thousand Oaks • New Delhi

© Mansoor A.F. Kazi 2003

First published 2003

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Inquiries concerning reproduction outside those terms should be sent to the publishers.



SAGE Publications Ltd
6 Bonhill Street
London EC2A 4PU

SAGE Publications Inc
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B-42, Panchsheel Enclave
Post Box 4109
New Delhi - 100 017

Library of Congress Control Number available

A catalogue record for this book is available from the
British Library

ISBN 0-7619-6995-0
ISBN 0-7619-6996-9 (pbk)

Contents

1. Introduction	1
2. Contemporary Perspectives in Practice Evaluation	10
3. The Realist Evaluation Paradigm for Practice	22
4. An Example of Qualitative Research Methods in Realist Evaluation	42
5. Quantitative Methods: Evaluation of Family Centres	62
6. A Realist Evaluation of Practice: the NSPCC's Shield Project and Intensive Analysis	95
7. The Shield Project: Extensive Analysis in Realist Evaluation	110
8. Patterns of What Works, For Whom and in What Contexts	137
9. The Contribution of Realist Evaluation for Practice	158
References	168
Index	176

List of Figures

Figure 3.1	The realist effectiveness cycle	29
Figure 5.1	Walpole case 11	68
Figure 5.2	Westtown case 10	69
Figure 5.3	Southgate case 20	70
Figure 6.1	K's outcome measures	105
Figure 6.2	K's outcome measures	105
Figure 6.3	K's outcome measures	106
Figure 6.4	H's outcome measures	107
Figure 8.1	The interaction of causal mechanisms in realist evaluation	148

List of Tables

Table 3.1	Change in majority of scored ADLs in Oakes Villa	38
Table 3.2	Destination of clients at discharge from Oakes Villa	38
Table 3.3	Result of DCC scores	40
Table 4.1	Schedule of repeated interviews	46
Table 4.2	Changes in outcomes matrix: template analysis of first interviews	47
Table 4.3	Changes in outcomes matrix: template analysis of second interviews	48
Table 4.4	Changes in enabling mechanisms matrix: template analysis of first interviews	50
Table 4.5	Changes in enabling mechanisms matrix: template analysis of second interviews	51
Table 4.6	Changes in disabling mechanisms matrix: template analysis of first interviews	52
Table 4.7	Changes in disabling mechanisms matrix: template analysis of second interviews	53
Table 4.8	Changes in content matrix: template analysis of first interviews	54
Table 4.9	Changes in content matrix: template analysis of second interviews	55
Table 4.10	Changes in context matrix: template analysis of first interviews	57
Table 4.11	Changes in context matrix: template analysis of second interviews	58
Table 5.1	Distribution of cases and numbers included in the evaluation	67
Table 5.2	Comparisons between the first and the last scores, and effect sizes	72
Table 5.3	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by age of child (in years)	73
Table 5.4	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by gender of child	73

Table 5.5	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by gender of the parent(s) working with the family centres	74
Table 5.6	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by the racial origin of the child	74
Table 5.7	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by profession of the referrer	75
Table 5.8	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by the reason for referral	76
Table 5.9	Parents' self-esteem (PCSEQ) outcomes cross-tabulated by type of service provided at the family centres	77
Table 5.10	Observation of Child Rating Scale (OCRS) outcomes cross-tabulated by age (in years)	77
Table 5.11	Observation of Child Rating Scale (OCRS) outcomes cross-tabulated by gender of the child	78
Table 5.12	Observation of Child Rating Scale (OCRS) outcomes cross-tabulated by gender of the parent(s) working with the family centres	78
Table 5.13	Observation of Child Rating Scale (OCRS) outcomes cross-tabulated by the racial origin of the child	79
Table 5.14	Observation of Child Rating Scale (OCRS) outcomes cross-tabulated by source of referral	79
Table 5.15	Observation of Child Rating Scale (OCRS) outcomes cross-tabulated by reasons for referral	79
Table 5.16	Type of service provided and OCRS outcomes	80
Table 5.17	Parent-child interaction (PCIS) outcomes cross-tabulated by age of child (in years)	81
Table 5.18	Parent-child interaction (PCIS) outcomes cross-tabulated by gender	81
Table 5.19	Parent-child interaction (PCIS) outcomes cross-tabulated by gender of the parent(s) working with the family centres	81
Table 5.20	Parent-child interaction (PCIS) outcomes cross-tabulated by the racial origin of the child	82
Table 5.21	Parent-child interaction (PCIS) outcomes cross-tabulated by sources of referral	83
Table 5.22	Parent-child interaction (PCIS) outcomes cross-tabulated by reasons for referral	83
Table 5.23	Parent-child interaction (PCIS) outcomes cross-tabulated by type of services provided by the family centres	83

Table 5.24	Parents'/carers' CSQ8 scores	84
Table 5.25	Parents'/carers' CSQ8-1 scores cross-tabulated by racial origin	85
Table 5.26	Parents'/carers' CSQ8-1 scores cross-tabulated by reasons for referral	85
Table 5.27	Parents'/carers' CSQ8-1 scores cross-tabulated by type of service provided	86
Table 5.28	Referrers' CSQ8 scores	87
Table 5.29	Referrers' CSQ8-1 scores cross-tabulated by racial origin	87
Table 5.30	Referrers' CSQ8-1 scores cross-tabulated by reasons for referral	88
Table 5.31	Referrers' CSQ8-1 scores cross-tabulated by type of service provided	88
Table 5.32	Patterns through odds ratios and relative risk	91
Table 7.1	Shield child BRIC outcome and effect size	119
Table 7.2	The effect sizes of changes in the parent, school and carer BRIC scores	119
Table 7.3	Self-concept effect size	120
Table 7.4	The effect size of changes in CAM and CAF scores	120
Table 7.5	Results of AI reliability tests and the effect size	121
Table 7.6	Effect size of change in Hare Self-Esteem Scores	122
Table 7.7	Effect size of change in A-COPE scores	123
Table 7.8	General case outcome: whether assessment was completed or not	124
Table 7.9	Engagement of the child/young person in the programme	125
Table 7.10	Engagement of either parent in the programme	125
Table 7.11	Assessment sessions held with the child/young person	126
Table 7.12	Assessment sessions held with either parent	126
Table 7.13	Number of days in assessment work and multi-disciplinary meetings held	127
Table 7.14	The extent to which the requirements of the assessment model were met	127
Table 7.15	Contexts in the circumstances of the children/young people ($n = 49$)	129
Table 7.16	Mechanisms and changes in the mechanisms ($n = 49$)	130
Appendix 7.1	Matrix of qualitative review templates against requirements of the model met and assessment completed	133

Table 8.1	Significant correlations between assessment completed and other outcomes, content variables and mechanisms	141
Table 8.2	Significant correlations between engagement of the child/young person in the programme and other outcomes, content variables and mechanisms	143
Table 8.3	Significant correlations between engagement of either parent in the programme and other outcomes, content variables and mechanisms	145
Table 8.4	Binary logistic regression with mechanisms	150
Table 8.5	Binary logistic regression in relation to contexts	151
Table 8.6	Binary logistic regression analysis of contexts impacting on the outcome of assessment completed (final step in the forward–conditional method)	153
Table 8.7	Binary logistic regression analysis of contexts impacting on the outcome of engagement of either parent (all three steps in the forward–conditional method)	153

This book is dedicated to my wife Shahida, my sons Jehan
and Shan, and to my daughter Shabby

Acknowledgements

I am indebted to Colin Robson and Nigel Parton for their helpful guidance throughout this enterprise. Particular thanks also to Ray Pawson and Nick Tilley for sending me a copy of their historic book that set the ball rolling. In the development of the realist evaluation paradigm, my thanks also to the colleagues from all over the world who have helped in the development of ideas and strategies – Ilmari Rostila, Mikko Mäntysaari, Tuija Lindquist, Riitta Haverinen, Marja-Riitta Kilponen, Björn Blom in Finland and Sweden; Bruce Thyer, Tony Tripodi, Jeane Anastas, Kevin Corcoran, Wayne Paris in the US. Above all, this book would not have been possible without the practitioners who took up the challenge to integrate the realist evaluation paradigm in their practice – the NSPCC Shield Project’s past and present workers, namely, Ann Ward, Valerie Charles, Martin Walker, Karen Higgins, Lisa Wilkinson, Judith Walsh, Diane Duckworth, Baseer Mir, Tina Zawada and former managers Sue Woolmore and Gerry Hudson. I am particularly grateful to Ann Ward for her work from the practitioner’s perspective (Chapter 6). For the adult rehabilitation programmes at Oakes Villa and Barton, my special thanks to Kathleen Firth, Avril Henson, Sandra Lickess, Robert Taylor and Ann Ward; for the family centres’ example, my thanks to all the family centre workers and, in particular, their manager Margaret Buckley and my co-evaluator in the project, Martin Manby. For the Lifeline example, many thanks to Lucy Spurling, Maggie Rogan and Claire Goodhind, and, in particular, Dawn Lawson and Claire Fraser for their help in the template analysis. For the development of the inferential statistics used in all the examples, I am indebted to the Calderdale and Kirklees Health Authority statistician John Varlow for his help in working out how the database should be set up, and what tests would be appropriate, as well as for calculating the odds ratios and relative risk in particular. For the references from Scandinavia and the translations, my thanks to Ilmari Rostila, Björn Blom, Stefan Morén and Anna-Lena Perdal. I am also grateful to colleagues in my other realist evaluation projects which have not been used as examples in this book, but which have helped in the development of the paradigm for practice, namely, Moray Council Children’s Services, Leeds Social Services’ Therapeutic Team, and Kirklees Education Social Work Service, to name a few. I am also grateful to many other colleagues in the Society for Social Work and Research, the American

Evaluation Association, and the European Evaluation Society, and to those who have attended the 'Evaluation for Practice' series of conferences, for the exchanges of ideas and experiences. My thanks also to other colleagues who have participated as researchers in the various projects, Shahida Kazi, Treena May and Robert Taylor. More recently, I am indebted to Finland's STAKES (The National Research and Development Centre for Welfare and Health, Helsinki) for my appointment as a consultant, and to over a hundred social workers at social services departments in Helsinki and Korso who are now beginning to integrate realist evaluation procedures outlined in this book into their practice. I am also grateful to colleagues at Seinajoki Polytechnic in Finland for their participation in developing realist evaluation. Finally, I am grateful to Sue Frost, Dean of Human and Health Sciences at this University, for her encouragement and guidance.

Confidentiality

In all of the examples, pseudonyms, letters of the alphabet or numbers are used to protect the identity of service users. In all the evaluations, informed consent was obtained from all participants. All the agencies where these evaluations took place have been named and identified with their permission.

Preface

This book is the culmination of work dating back to 1990, when I was Education Officer (Pupil Welfare) and head of a newly-formed Performance Review Team at the Rochdale Local Education Authority near Manchester in England. Having moved to the University of Huddersfield, I was able to develop evaluation strategies with Joe Wilson at the Kirklees Education Social Work Service, Kathleen Firth at Oakes Villa Rehabilitation Unit, and Michelle Hayles at the West Yorkshire Probation Service. The extensive use of single-case evaluation led to contacts with Bruce Thyer and other colleagues from the Society for Social Work and Research (SSWR) in the USA. By 1995, the Centre for Evaluation Studies (and later also the Centre for Applied Childhood Studies with Nigel Parton) at the University of Huddersfield began a series of 'Evaluation for Practice' international conferences, and I was also influenced by the conferences of the SSWR, the European Evaluation Society (EES) and the American Evaluation Association (AEA). Having used outcome studies extensively, it became apparent that this was a foundation that needed to be built upon. How could we determine the content of the services that were found to be effective, and the contexts in which they were effective? I was influenced by Juliet Cheetham and other colleagues who were then at Stirling, and began to apply the pragmatic approaches. At the 1997 EES conference in Stockholm, I presented a paper in the same session as Ray Pawson and Nick Tilley. Their paper on realist evaluation appeared to provide some answers, but it did not become clear until they sent me their newly-published book. Having worked with social work, health, education and probation practitioners and developed the empirical practice approach of integrating research methods into practice, I began to continue with this emphasis, and began to integrate the realist effectiveness cycle where it was possible. This book presents the findings from the first several evaluations where the realist paradigm has been applied, based on my PhD thesis.

1

Introduction

This publication aims to outline and develop a framework for the application of the new, emerging realist paradigm in evaluation research in practice, and applies it to social work practice in particular and to the practice of human service programmes in general. This paradigm is reported to have the potential for an evaluation strategy that not only systematically tracks outcomes, but also the mechanisms that produce the outcomes, the contexts in which these mechanisms are triggered, and the content of the interventions (Kazi, 2000a; Pawson and Tilley, 1997b). According to realism, human service programmes introduce interventions that interact with the other causal mechanisms in the contexts of service users to produce outcomes. The purpose of realist evaluation is reportedly to investigate ‘what works best, for whom, and under what circumstances’ (Robson, 2002: 39). However, to date the realist paradigm largely remains at the level of a philosophy of science (Bhaskar, 1998; Sayer, 2000), and as a manifesto for evaluation research (Pawson and Tilley, 1997b; Robson, 2002). As the title suggests, this publication attempts to develop an approach for the integration of realist evaluation into the practice of human services, and the methodologies that can be used for realist evaluation. Practice examples are provided from a number of studies to substantiate the potential contribution of realist evaluation for practice.

To date, there is a dearth of published examples of realist evaluations in any area of human service practice which can demonstrate the utility of the paradigm or the processes and methodologies that can be used to actually achieve an investigation of ‘what works best, for whom, and under what circumstances’. The paucity of examples is exemplified by the fact that the advocates of realism tend to use the same few examples as illustrations (e.g., Pawson and Tilley’s, 1997b example of crime prevention measures which also appears in Robson, 2002 and Sayer, 2000). This publication attempts to make a contribution by providing real examples and helping to develop ways in which this philosophy of science could be translated into an evaluation paradigm for practice. In particular, this publication develops realism as a paradigm for practitioner-evaluators, with the development of a realist effectiveness cycle which can be integrated into the practice of

human services in a partnership between internal evaluation (that is, what the agencies themselves do) and external evaluation (that is, evaluations carried out by academics and external consultants). In this way, the findings from evaluation can be directly utilised to influence the future development of programmes.

Evaluation for Practice

The pressures on social work practice to demonstrate effectiveness have continued to grow in the last two decades. The pressures from changes in the legal and societal context mean that social work is no longer taken for granted and that its worth has to be demonstrated (Parton, 1994). In Britain, the Children Act 1989 and the Community Care Act 1990 both included requirements for planning in response to need, and reviewing progress. The purchaser–provider split, the growth of the voluntary and private sectors alongside the public sector, and the introduction of competition for contracts also made monitoring and evaluation more central in social work practice. The current British government has continued this trend with an emphasis on ‘league tables’ which rank health and social services according to performance. There is a growing emphasis on evidence-based practice and clinical effectiveness (Powell, 2002). The resources are finite, and yet the social needs are complex and in a state of flux. Evaluation research is one way to make social programmes accountable and to enable politicians, agencies and practitioners to make hard choices in the allocation of scarce resources. The analysis thus far has concentrated on the question of accountability and demonstrating the merit and worth of social work. There is another dimension to this – the need to develop and improve the content of social work practice itself, so that it is better able to meet the needs of its clients and the wider society.

According to Mark, Henry and Julnes:

Evaluation assists sensemaking about policies and programmes through the conduct of systematic inquiry that describes and explains the policies’ and programmes’ operations, effects, justifications, and social implications. The ultimate goal of evaluation is social betterment, to which evaluation can contribute by assisting democratic institutions to better select, oversee, improve, and make sense of social programmes and policies. (2000: 9)

Robson adds: ‘Evaluation is often concerned not only with assessing worth or value but also with seeking to assist in the improvement of whatever is being evaluated’ (1993: 175). Therefore, there are two main purposes of evaluation research – providing evidence of the merit and worth of social work practice, and striving to improve practice itself to respond to the changing needs and contexts, for the betterment of society. Whether emphasis is placed on one or the other of these purposes may depend on the paradigmatic influences that are inherent in the effectiveness inquirer’s

activities. For example, from a critical theorist perspective, Shaw adds another purpose in evaluation for practice: 'Evaluating in practice is not limited to determining whether social work is effective, but must be a means of empowerment and social change' (1996: 189).

The epistemological debates in philosophy and in other sciences are also reflected in social work research, with perspectives ranging from the 'empiricist' view that effectiveness can be demonstrated through empirical evidence of effects, to the 'humanist' or interpretivist position that effectiveness can be demonstrated through the subjective perspectives and the meanings attached to such perspectives (as described in Shaw, 1996: 21).

In response to these developments in the philosophies of science, there has been a growth in research methods' textbooks and other publications addressing the need for social work to demonstrate its effectiveness (Newburn, 2001). Most of the authors have tended to be university-based, but these publications also reflect a developing partnership between academics and social work practitioners. For example, Macdonald (1996) is one of a number of publications on effectiveness from Barnardos – a children's charity and a voluntary social work agency; and Everitt and Hardiker (1996) is a British Association of Social Workers' publication. Kazi (1998a) and Fuller and Petch (1995) directly address practitioner research, and Shaw (1996) has a number of examples from practice. These and other publications are contributing to the development of effectiveness strategies that can be applied to human service practice, by both practitioners and researchers.

This book also seeks to make a modest contribution to the development of evaluation research in contemporary practice in social work, health and other human services. The book begins properly in Chapter 2 with a critical analysis of contemporary paradigmatic approaches to the evaluation of practice, including the extent to which each approach is able to 'capture' the breadth and depth of the effectiveness of practice. In the main (but not exclusively), the recent publications referred to above reflect the influences of some of the main paradigms from the philosophies of science. Each of these approaches can be credited with the contribution it has made, and continues to make, to various aspects of practice effectiveness – this can be substantiated through an analysis of the ontology, epistemology and methodologies (Guba, 1990) associated with each paradigm. At the same time, each of these paradigms also has its limitations, and this critical analysis will attempt to show the extent to which any one of these paradigms can address the complexities of practice effectiveness.

Paradigms and Influences

The term *paradigm* is used a great deal in this book, and therefore it is important to clarify what this means from the start:

Close historical investigation of a given speciality at a given time discloses a set of recurrent and quasi-standard illustrations of various theories in their conceptual, observational, and instrumental applications. These are the community's paradigms, revealed in its textbooks, lectures, and laboratory exercises. By studying them and by practising with them, the members of the corresponding community learn their trade. (Kuhn, 1970: 43)

The recent publications in social work effectiveness research (reviewed in more depth in Chapter 2) encompass the main contemporary paradigms, as Kuhn suggests. However, like most of the terms used in this book, precision in definitions is not possible, as concepts tend to be used in different ways, and the definitions shift according to the way a term is used. Kuhn is credited with the introduction of the notion of the paradigm, but Masterman (1970) notes that he used the term in 21 different ways. In this book, the term *paradigm* is used in the sense of a 'set of beliefs about the nature of the world and the individual's place in it' (Mark, 1996: 400). In an operational sense, paradigms are characterised by the inquirer's world view, or his/her outlook with respect to the existence of reality, the theory of knowledge, and the way one conducts an inquiry. In the practice of evaluation, discrete boundaries cannot be drawn for each paradigm (Kazi, 2000a; Shaw, 1996; Trinder, 1996), as the evaluator is likely to cross many a boundary. Nevertheless, paradigmatic influences can determine the selection of evaluation questions and the selection of research methods to deal with those evaluation questions.

The Realist Evaluation Paradigm

Realism is ... a common-sense ontology in the sense that it takes seriously the existence of the things, structures and mechanisms revealed by the sciences at different levels of reality ... the task of science is precisely to explain 'facts' in terms of more fundamental structures, and in the process it may reveal some of these 'facts' ... to be, in part, illusions ... we may not yet, and never finally, know whether it is true or false. (Outhwaite, 1987: 19–20)

The term *fallibilistic realism* was first suggested by Donald Campbell in a personal communication (Manicas and Secord, 1983); and it is also used by Anastas and MacDonald (1994) (and more recently by Anastas, 1999) who were the first to introduce this perspective in social work effectiveness research. However, if we include texts in the Finnish language (see Rostila and Kazi, 2001), then Professor Mikko Mäntysaari (University of Jyväskylä, Finland) wrote about realism in relation to social work a few years before Anastas and MacDonald. This perspective is also known by other terms, such as *scientific realism*, *transcendental realism*, *referential realism* or generally as a *realist* view of science or even as *post-positivism* (Fraser et al., 1991; Phillips, 1990). Archer (1998), Bhaskar (1998) and Robson (2002) prefer the term *critical realism*, emphasising realism's critical role in social research. In the application to evaluation for practice, this author prefers the term *realist evaluation*, which is similar to *realistic evaluation* (Pawson and Tilley, 1997b).

However, *realistic evaluation* implies a tendency whereas *realist evaluation* is a more emphatic description of this new paradigm in evaluation research.

Realist evaluation is based on the work of the philosophers Roy Bhaskar (1997, 1998) and Rom Harre (1984). Mark, Henry and Julnes (2000) and Pawson and Tilley (1997a, 1997b) have developed the realist paradigm as a legitimate evaluation research perspective in its own right. Anastas (1999), Kazi (1998a, 1999, 2000a, and 2000b) and Rostila (2000, 2001) attempt to apply this perspective in the evaluation of social work practice; and in other human services, Porter and Ryan (1996), Tolson (1999) and Wainwright (1997) apply realism in health services.

Realist evaluation seeks to evaluate practice within the realities of society. Practice takes place in an open system that consists of a constellation of interconnected structures, mechanisms and contexts. Realism aims to address all the significant variables involved in social work practice, through a realist effectiveness cycle which links the models of intervention with the circumstances in which practice takes place.

Realist evaluation research is about improving the construction of models and, therefore, about improving the content of the practice itself. Evidence from data gathering is used to target and adjust the content of the programme in such a way that it can have a generative impact on pre-existing mechanisms and contexts, and help to bring about the desired changes. Objectivity lies not just in the use of outcome measures, but in the extent to which the model is analogous with reality. At each cycle, a better approximation of reality is obtained, as compared with the previous cycle. In this way, realism addresses all the dimensions and questions of effectiveness of practice, including contexts, the perceptions of all involved, ethics and values, and the content of practice. The multi-method data gathering addresses the questions of what actually works, for whom and in what contexts. (Kazi, 2000c: 317)

Underdevelopment of Realist Evaluation Methodology

The above definition provides a summary of this author's development of the realist paradigm in the evaluation of practice in human service programmes. However, a contention of this book is that whilst realism is developed as a philosophy of science, at the level of methodology this paradigm is relatively underdeveloped at this stage. For example, at the sixth annual conference of the International Association for Critical Realism (IACR) in August 2002, one of the three main themes was 'research using realism':

What constitutes critical realist empirical research? In what ways does a critical realist perspective influence or facilitate substantive research? We are particularly interested in papers providing answers to these and other questions in the best possible way – by reporting the results of substantive research undertaken from a realist perspective. (p. 1, http://www.criticalrealism.demon.co.uk/iacr/conference_2002.html)

The fact that this is still a major theme suggests that these issues are far from being resolved at the level of realist research designs generally, and the same is true for evaluation research. There are no complete published

realist evaluations of human services; and to date, no account of dedicated methodologies that can be applied to investigate what works, for whom and in what circumstances. For example, Pawson and Tilley (1997b) outline the rules and framework for a realist evaluation, but they fall short of specifying the methodologies for investigating what they call mechanism–context–outcome configurations. Robson (2000) has a helpful section on mechanisms and how they influence programme design and evaluations, but again does not explicitly state the methodologies that may be appropriate for a realist evaluation. Robson (2002) does suggest that inferential statistics may be used, as well as qualitative methods as specified by Miles and Huberman (1994); however, no complete examples of realist evaluation are provided.

As Tolson explains, realist evaluation

is an applied form of research which lends itself to the process of practice innovation through its contextual sensitivity. Accordingly it is complex and its methodological rules are still emerging. The apparent complexity will undoubtedly ease as this type of evaluation research matures and its practice is documented. (1999: 389)

Realism is methodological-pluralist, but the methods it can draw upon were developed either within the empirical or interpretivist paradigms that may not have the same ontological depth as realism. These issues are considered in Chapter 3, and then the rest of this publication provides examples of the application of realism in the evaluation of human services – with the aim of easing the complexity of realist evaluation, as indicated by Tolson.

Issues of Implementation and this Book's Contribution

Realism transcends the qualitative and quantitative divide, or the epistemological divide between empirical and interpretivist approaches (Mark, Henry and Julnes, 2000; Pawson and Tilley, 1997b). Nevertheless, realists continue to be influenced by these debates and express a preference for either of these approaches, at the methodological level, even within the stated methodological-pluralist approach in realism. For example, Sayer (2000) advocates intensive research (which he defines in qualitative terms), as the only way of achieving an explanatory critical realist inquiry. On the other hand, Lawson (1998) emphasises the need to investigate demi-regularities, and thereby advocates extensive research, with an emphasis on empirical methods. These preferences are within the paradigm of realism, as it is expected that the selection of evaluation questions, and how the research is conducted, depend upon the theoretical and methodological preferences of the inquirer. However, the preferences expressed by these two authors are of crucial importance at this stage when realism is still an emerging evaluation research perspective, and when there is a need to develop this paradigm further in its actual application to the evaluation

of human services. Realism is relatively underdeveloped at the level of methodology at this stage; and therefore, the question remains – can you go further from ‘what works’ to ‘what works, for whom and in what contexts’? These questions are addressed in the subsequent chapters of this book, with real examples of realist evaluation to contribute to the development of appropriate methodologies and to help ease the apparent complexities of realist evaluation.

Chapter 2 attempts to categorise the main evaluation research perspectives, and the contribution to practice evaluation made by each perspective. The empirical practice approach with a focus on outcomes provides a foundation for evaluation, and the other perspectives (that is, interpretivist and pragmatist) add building blocks to the process of evaluation. However, these approaches remain at either the ‘black box’ (outcomes only) or ‘grey box’ (outcomes with some components of process) levels. The ‘white’ (or preferably, ‘clear’) box evaluations are the potential contribution of the realist evaluation perspective.

Chapter 3 outlines some key concepts from the realist perspective that are relevant for the practice of evaluation. The outcomes of a programme can be understood in relation to the causal mechanisms that produce them, and the contexts in which they are triggered. Investigations of these mechanism–context–outcome configurations enable an account of the circumstances in which a programme may be more successful, and the circumstances in which it may be less successful. A framework for practice, or ‘realist effectiveness cycle’ is proposed that enables an integration of realist evaluation procedures into a programme’s practice, and establishes a direct link between practice and evaluation in order to improve practice. The chapter provides an example from adult rehabilitation services where this cycle was beginning to be integrated, with the use of outcome measures in daily practice. When analysed with the other patient information in the agency’s records, the ‘black box’ began to turn ‘greyer’, indicating the utility of realist perspective in encouraging a search for explanations beyond appearances.

Chapter 4 provides an example of the use of qualitative methods in realist evaluation. Five sets of repeated interviews from a project with the drug-using community are used to demonstrate the identification of mechanisms, contexts, content, and outcomes from the service users’ perspectives. Template analysis was used to identify the patterns that emerged from the data. As an example of a ‘grey box’ study, a limitation was that outcomes were not systematically tracked, and therefore this example also remained at the beginning stages of the realist effectiveness cycle. The example also demonstrates that the use of a single method (whether quantitative or qualitative) may enable the identification of mechanism–context–outcome patterns, but tends to fall short of establishing the causal factors that may be responsible for change.