

Keynote Address
Thirteenth National Symposium on Doctoral Research in Social Work
College of Social Work
April 6, 2001

Realist Evaluation for Practice

Mansoor A.F. Kazi
Reader and Director of Centre for Evaluation Studies
The University of Huddersfield
Queensgate, Huddersfield
HD13DH, England

This paper outlines the new, emerging realist paradigm in evaluation research, and applies it to social work practice. This paradigm has the potential for a 'white box' evaluation 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 (or the generative mechanisms introduced by a programme). Two examples are provided, both studies with an extensive use of single-subject designs by practitioners within a realist paradigm.

Social work interventions usually take place at the interface of the individual and social, where multiple factors and influences are continuously at work (Cheetham et al., 1992). The extent to which these complexities are addressed in practice evaluation, depend upon a) the paradigmatic perspective of the researcher, and b) the extent to which the particular perspective enables the researcher to address these complexities. Based on a review of literature (Kazi 1999, 2000), the main contemporary perspectives in British social work practice research could be identified as:

- 1) Empirical practice (Thyer 1998, Sheldon & Chilvers 2000) that emphasises evaluation activities based on outcomes and concentrates almost exclusively on the effects of practice as defined in terms of measurable outcomes. Future successes cannot be guaranteed not only because of the inadequate descriptions of content which make replication difficult, but also because typically there is no analysis of contexts which are inherently unpredictable.
- 2) The interpretivist approaches including several epistemologies such as critical theory, feminist evaluation and social constructionism (Taylor & White 2000,

Parton & O'Byrne 2000). However, these perspectives tend to be suspicious of outcome-based methodologies, and therefore their focus tends to be one-sided in capturing the dimensions of practice.

3) The pragmatic, methodological-pluralist approach (Cheetham 1998) recognises the limitations of both empirical practice and interpretivist approaches, and attempts to provide a perspective that goes beyond the consideration of either outcomes or interpretivist insights. However, the pragmatic focus means that concentration is on what is seen to be desirable and appropriate at any time.

Each of the above three perspectives has its limitations, based on emphasis on one or the other element of the complexities of practice; at the same time, each has an important role to play in addressing these complexities, and each sets out to achieve this goal in its own way. Another way of categorising the evaluation strategies is to consider the three 'boxes' of evaluation. Adapting Michael Scriven's terminology of 'black', 'grey' and 'white box' evaluations (Scriven 1994), 'black box' evaluation is where the researcher concentrates on evaluating a programme's effects, without addressing the components that make up the programme. Such research is crucially important, and stands in its own right; this is the role of much empirical practice research. 'Grey box' evaluation is where the components of a programme are discerned, but their inner workings or principles of operation are not fully revealed; this is the contribution of much pragmatist and interpretivist research.

Another post-positivist perspective is emerging in social research, that of scientific or critical realism (Bhaskar 1998, Pawson & Tilley 1997b). Realism attempts Scriven's 'white box' evaluation, which not only addresses the effects, but also the inner workings and operations of the components of a programme and how they are connected. This new, emerging paradigm appears to have the answers for dealing with the apparent limitations of these other contemporary perspectives, but as yet there is no report of a completed study in social work or health where this perspective has been applied. The studies described in this paper make a contribution in this regard.

Duguid and Pawson (1998) refer to the three 'boxes', when they describe their efforts to make the 'black' box 'greyer'; and Kazi (2000a) presents an account of the author's evaluation activities as classified into the three boxes. When the purposes of evaluation are categorised within these three types of approaches, or 'boxes', empirical practice which concentrates on the evaluation of effects (e.g. through single-case evaluation or through comparison group designs) forms the basis in each type of box. Other methods are added to address the wider questions, but it is assumed that empirical practice, as in the form of the 'black box', remains. In this sense, the other perspectives add building blocks to the process. It is important to recognise what each perspective contributes to the needs of practice evaluation, and to adopt an inclusive approach where one perspective does not have to destroy another in order to create a niche for itself.

The realist evaluation paradigm

This perspective is known by a number of different terms, e.g. *fallibilistic realism* in Anastas (1999), or *post-positivism* (Fraser et al 1991, Phillips 1990). This author prefers the term “*critical realism*”, in keeping with Bhaskar (1998) and Archer (1998): “...usually not all (of our environment) is revealed to consciousness and sometimes that is because it is shaped outside of our conscious awareness. Revealing the latter is what makes for critical realism and its emancipatory potential...” (Archer 1998, p. 1999). In the application to evaluation for practice, this author prefers the term “*realist evaluation*”, which is similar to *realistic evaluation* (Pawson & Tilley 1997b). However, *realistic evaluation* implies a tendency whereas *realist evaluation* is a more emphatic description of this new paradigm in evaluation research.

The critical realist perspective is inclusive of all three perspectives (empirical practice, interpretivist and pragmatic approaches) and shares the same realist ontology with them. However, critical realism goes further than the other paradigms in recognising that the world is an open system or a constellation of structures, mechanisms and contexts. At the level of ontology, the recognition that reality is stratified necessitates the rejection of the other world views that rely upon surface sense data. The critiques of positivism (e.g. interpretivist approaches) reject its empirical epistemology whilst at the same time retaining the flat, unstratified ontology. Critical realism, on the other hand, rejects the flat ontology of all of these other perspectives, and distinguishes between the *real*, the *actual* and the *empirical* (Sayer 2000). The *real* exists regardless of our understanding of it, and constitutes the realm of objects, their structures and powers. The *actual* refers to what happens if and when these powers are activated. The *empirical* is the domain of experience that can refer to either the real or the actual. Some structures may not be observable, but can be inferred by reference to observable effects that can only be explained as the products of such entities. Critical realism’s stratified ontology is not only in contrast to the flat ontology of other perspectives, but also includes the concept of emergence—i.e., that the world is characterised by emergence or that the stratification of structures continually gives rise to new and emerging phenomena. Emergence is inherent in the concept *structures*, which suggests “a set of internally related elements whose causal powers, when combined, are emergent from those of their constituents” (Sayer 2000, p. 14).

“The aim is not to cover a phenomenon under a generalisation (this metal expands when heated because all metals do) but to identify a factor responsible for it, that helped produce, or at least facilitated, it” (Lawson 1998, p. 156). This process, known as *retroduction*, is based on critical realism’s stratified ontology and differentiates the realist paradigm from the others. Unlike these others, the critical realist inquirer—who can never operate under conditions of complete closure in an open system—will seek to develop a theoretical understanding of the nature of the structure such that he/she can explain the causal mechanisms, and the conditions under which certain outcomes will or will not be realised. Causal powers do not reside in the events or the behaviours of particular objects, variables or individuals, but in the social relations and organisational structures which constitute the open system. One action leads to another because of the actions’ accepted

place in the whole. Persons are complex particulars and the events of interest---e.g. programme outcomes associated with human services---are the result of complex transactions of many different kinds of structures at many different levels and cannot be explained simply in terms of a causal link between events at the surface.

In the critical realist world view, social work programme outcomes cannot be explained in isolation; rather, they can only be explained in the sense of a mechanism that is introduced to effect change in a constellation of other mechanisms and structures, embedded in the context of pre-existing historical, economic, cultural, social and other conditions. "On the transcendental realist view of science, then, its essence lies in the movement at any one level from knowledge of manifest phenomena to knowledge produced by means of antecedent knowledge, of the structures that generate them" (Bhaskar and Lawson 1998, p. 5). In this way, effectiveness of the programme is apprehended with an explanation of why the outcomes developed as they did, and how the programme was able to react to the other underlying mechanisms, and in what contexts. This analysis provides not only evidence of effectiveness, but also an explanation that helps to develop and to improve both the content and the targeting of future programmes.

Application of realist evaluation for practice

Based on the analysis of data including outcomes, mechanisms and contexts, programmes are developed as models of intervention targeted to achieve the desired outcomes. A multi-method strategy is applied to test the extent to which these models of intervention are analogous with reality; and the data collection and analysis directly contribute to the further development of these programmes of intervention as well as their future targeting within a realist effectiveness cycle (Figure 1). A cycle is selected as, unlike natural sciences, "instead of running straight ahead in pursuit of new knowledge, they (social sciences) move around in small circles and spend a lot of time re-inspecting the starting block" (Outhwaite 1998, p. 290). Therefore, the starting point in figure 1 is theory that includes propositions on how the mechanisms introduced by a programme into pre-existing contexts can generate outcomes. This entails theoretical analysis of mechanisms, contexts and expected outcomes, using a logic of analogy and metaphor (Bhaskar and Lawson 1998). Theoretical explanations are characteristically analogical, e.g. scientists began looking for a virus for mad cow disease as previous ailments in cattle tended to be caused by a virus (Lawson 1998). In the same way, in evidence-based practice social workers' theoretical constructs would be based on what is known about their particular areas of work. The practitioner would draw upon prior knowledge of causal mechanisms which accounted for the effectiveness or otherwise of models of intervention in particular contexts. The model will include assessments of personal, social and environmental difficulties; a programme of intervention designed to help through supportive, rehabilitative, protective or corrective action, and expectations of changes.

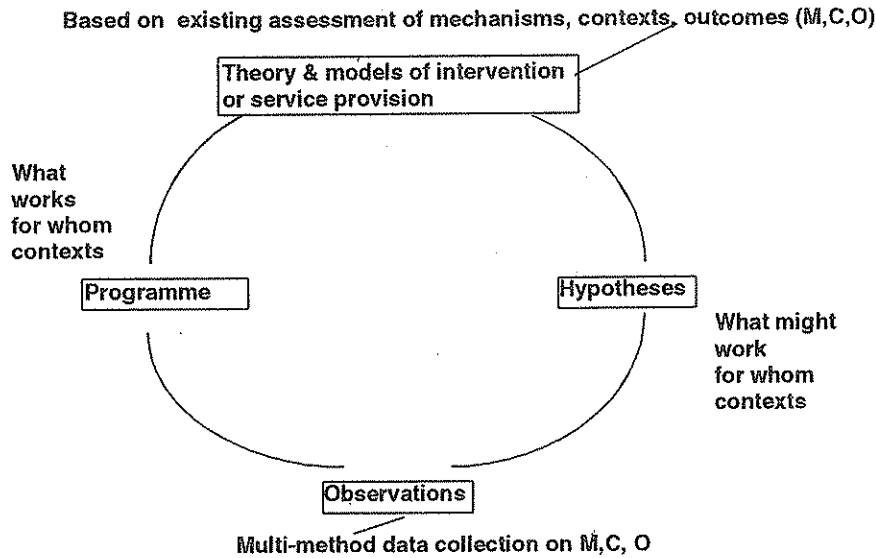


Figure 1: The realist effectiveness cycle
 Adapted from Pawson and Tilley (1997b)
 Source: Kazi (1998, 1999)

The second step on the cycle consists of hypotheses based on realist abstraction. The hypotheses would typically address the following questions:

1. who is the programme targeted at,
2. what changes will be brought about,
3. what contexts impinge on this,
4. what social, cultural and other mechanisms in the pre-existing environment would enable these changes, and which ones may disable the programme (i.e., countervailing mechanisms); and
5. what other potential mechanisms may be generated by this programme in the process of change.

The next step on the cycle is the selection of appropriate methods of data collection to help return to the concrete---and here, realists are committed methodological-pluralists and do not rule out anything that addresses real entities. The realist inquirer would identify the evaluation research methods that can address the questions raised by the theory and the hypotheses, and that can also provide data on the theoretical propositions, the identified mechanisms, and the identified outcomes of the programme. It is here that a plausible connection can be made between the social work model and its likeness with reality---or, to put it another way, provide evidence of the model's ability to change reality.

Further on, we turn to the actual programme, in order to make it more specific as an intervention of social work practice. This specificity is based on the findings from research methods to date, i.e. an investigation, so far, of what works for whom and in what contexts, to target the programme better, and to improve its content to meet the needs of the users it

is actually aimed at. Therefore, specificity is also desirable in the explanation of the contexts in which the programme operates and in which the service recipients are embedded. What works may differ according to the different contexts of the service users. The programme may be directed at one person, a family, a group, a community, or an entire population---it will be based on explanations, so far, of the role of particular mechanisms embedded in particular contexts to ensure that the programme has maximum impact.

Next, but not finally, we return to the theory---not finally, because the cycle continues. The theory (including assessment) is developed further, the hypotheses are based on explanatory evidence thus far, the data collection methods are developed and applied more appropriately, and the programme is developed accordingly, and returning to theory, and so on. The social worker has his/her own models of practice that he/she follows in making assessments as well as service delivery to a client or a client group. The realist effectiveness cycle enables a dialectical relationship between this model and the realities of practice, which enables the refinement and development of this model based on the realities of practice. It is based on evidence that goes into the depth of fluid contexts rather than remaining at the surface. This evidence can also be used as evidence of effectiveness, and to make judgements about the merit and worth of practice, but such use is a by-product---the real purpose is the development (and improvement) of the models of practice.

Examples of realist evaluation

The Centre for Evaluation Studies is undertaking a number of 'white' box evaluations at the present time, building on the progress made in the previous studies undertaken by this author with social work and health agencies, particularly with regard to the empirical practice approach to evaluation of social work practice (Kazi 1998, 2000a). This provides the foundation, upon which the critical realist approach is being built, to provide a more complete evaluation that tries to encompass the complexities of practice.

The examples of studies used in this paper contribute to the development of methodologies that are appropriate for a realist evaluation. An explanatory account will seek to penetrate the surface, to investigate the mechanism-context-outcome configurations, and to explain how the programme's causal mechanisms interacted with the other causal mechanisms in the circumstances of the service users, and the conditions or contexts in which they were triggered. This is the challenge of realist evaluation for practice, and the foregoing studies seek to respond to this challenge

Study 1: Integrating single-subject designs in family centres

The first example is a study involving the use of single-subject designs in the evaluation of five family centres in Kirklees local authority in West Yorkshire, England. The family centres provide day centre facilities for young children and their families in order to enhance family functioning and appropriate child development. The practitioners themselves selected the outcome measures, in consultation with the authors of the project report (Kazi, Manby and Buckley 2001). In a number of sessions, three main instruments were constructed: the Observation of Child Rating Scale (OCRS), Parent/Carer-Child Interaction Scale (PCIS),

and Parents'/Carers' Self-Esteem Questionnaire (PCSEQ). The OCRS measures the child's developmental milestones, including behaviour and play; the PCIS includes practical care, parental responsibility, management of child's behaviour, communication and play; and the PCSEQ measures the parent/carer's feelings and self-concept as a parent. Test-retest reliability for all three measures taken together was found to be .9156 (Kazi, Manby & Buckley 2001). In addition, the Client Satisfaction Questionnaire (CSQ8, Fischer & Corcoran 1994) was selected as a measure for parents, and adapted as a measure for referrers who were usually health visitors or field social workers such as those involved in child protection work. It was agreed that the OCRS, PCIS and PCSEQ would be used monthly, and that the CSQ8 would be used every two months with parents/carers and with the referring agencies.

As in the other single-case evaluation studies based in this region of England (see Kazi 1998), the family centre workers responded very positively. During the six-month period of evaluation, 29 family centre workers used single-case designs with 155 (or 78% of the total) children and their families in all five family centres. In all of these cases, the outcome measures were used more than once. However, the totals for each of the outcome measures fall short of 155 as not all the outcome measures were used repeatedly in every case (table 1). The computer software programme SPSS (Foster 1998) was used to analysis the data from the 155 cases. The extent to which progress was made in each case was determined by a comparison between the first and the last scores against each measure used. Table 1 indicates overall progress in all cases by a comparison of the first and the last scores in all cases across the three outcome measures.

Table 1: Comparisons between the first and the last scores

Change in outcomes	PCSEQ: no. of cases	PCSEQ: as % of cases (n=117)	OCRS: no. of cases	OCRS: as % of cases (n=126)	PCIS: no. of cases	PCIS: as % of cases (n=117)
No change	18	15.4	11	8.8	17	14.5
Improved	52	44.4	87	69.0	63	53.9
Deteriorated	47	40.2	28	22.2	37	31.6
Totals	117	100	126	100	117	100

The comparison between the first and the last scores indicate that as the interventions progressed over several months, there were significant improvements for the OCRS score and the PCIS score, but the results of the PCSEQ were more mixed. These findings were corroborated by high levels of satisfaction indicated by both parents/carers and the referring social workers with the use of CSQ8. It can be concluded from the above data analysis that in the six-month period, the family centres were largely effective across all the outcome measures used. At this stage, these are the conclusions of a 'black' box type of evaluation. Sub-group analysis can help to turn the 'black' box 'greyer' (Duguid & Pawson 1998) by identifying useful patterns in the data, and revealing some potential mechanism-context outcome configurations which otherwise may remain hidden. As

Lawson explains, "The significance of patterns collected under the heading of demi-regularities) usually turns upon comparisons, and in particular, upon differences: between men and women; old and young; events or states of affairs...we notice the effects of sets of structures through detecting relatively systematic differences in the outcomes of prima facie comparable types of activities..." (Lawson 1998, p. 153).

The outcomes were cross tabulated by age, gender, race, source of referral, and type of service provided, in order to identify the mechanisms that produced the outcomes, and the contexts in which these mechanisms were triggered. It was found, for example, that parents of younger children fared better in improving their self-esteem. It may be that the programme is more intensive with the younger children, and there is greater voluntary contact with the parents who may tend to stay for a while when bringing their babies and toddlers to the family centre, than with the parents of the children aged four years. There were also indications that the role of family centres as generative mechanisms in improving both the parents' self-esteem and parent-child interaction outcomes were not triggered in the same way for non-white as they were for white parents. There may be countervailing mechanisms associated with greater levels of social deprivation and discrimination with the specific categories of racial groups, or with cultural norms that were not taken into account in the family relationships. However, the mechanism-context configurations associated with racial origin did not influence the OCRS outcomes in the same way, as those outcomes relate to the development of the child rather than in relation to the parent. This suggests that these countervailing mechanisms are more likely to exist in relation to non-white parents than in relation to non-white children.

Proportionally greater improvements in both OCRS and PCIS were achieved in the cases where 'supporting the child's development and educational needs' was the type of service provided, and this may be because of more structured models of intervention acting as generating mechanisms directly with children. There were proportionally lower PCIS outcomes where 'parenting skills' was the type of service provided, even though parent-child interaction was an important part of parenting skills. However, the PCIS outcomes with regard to parents involved in 'group work' were very high, suggesting that work with parents in groups produced better outcomes with regard to parent-child interaction than individual work on parenting skills. This may be due to causal mechanisms triggered in the socialisation of parents, or in the way the intervention acted as a generative mechanism in the process of group work. As Duguid and Pawson (1998) found in their analysis, "it is not programmes that work but their capacity to offer resources that allow participants the choice of making them work" (p. 492).

The patterns in the data observed here provide only potential explanations and therefore further investigation is required to identify the mechanism-context-outcome configurations associated with these types of problems. One of the limitations is that, although both intensive and extensive research strategies have been used, the methods are entirely quantitative. In fact, the findings in this chapter corroborate Sayer's (2000) arguments that such extensive methods, at best only identify mechanisms that require further investigation through intensive qualitative research. The indications are that, the

use of quantitative strategies alone is not sufficient to determine the causal conditions, and the contexts in which they are triggered, with regard to human service programmes. However, they may be the first steps towards a retroductive analysis as indicated in the next example; and they may enable a retrospective analysis of quantitative data within a realist framework as in the example described in the foregoing.

Study 2: The Shield Project

Study 2 is based on the evaluation of a social work team that works with children and young people who sexually harm others. The Shield project is a collaborative venture involving the National Society for the Prevention of Cruelty to Children (NSPCC) and Kirklees Social Services Department in West Yorkshire. The project consists of a team of social workers who provide assessment and treatment facilities for children and young people who sexually harm others (Kazi & Ward 2001). The realist evaluation framework applied in this study enables the systematic tracking of outcomes, mechanisms, contexts, and the content of the interventions in each case, in order to achieve an analysis of what works, for whom and in what contexts (Kazi 1998).

For each case in this particular study, there is a qualitative analysis of the data, based on individual session records, supervision notes, plans, reviews and reports, as described above; and there is also quantitative data from the use of single-case evaluation (see Kazi & Ward 2001 for a more detailed account of the recording systems). Cross tabulation of outcomes and content of interventions in this particular application is based on the use of single-case evaluation, or the use of single-subject designs by practitioners to track client progress systematically or to evaluate the effectiveness of their interventions of (Kazi 1998 and 1998a; Bloom, Fischer and Orme 1999). Several standardised measures were selected by the Shield staff team from Fischer & Corcoran (1994), and a measure for victim empathy was also created and is under-going reliability tests at the time of writing. As the measures selected were of American origin, a few minor changes in wording were made in some of the measures, to make the language more specific to local requirements. These changes may have a minor affect on the reported alphas indicated in the original descriptions of the measures. In particular, the Behaviour Rating Index for Children (BRIC) was adopted as a basic measure for use in all cases where repeated measurement was possible; all of the other measures were in relation to outcomes specific to particular clients. The BRIC measures children's behaviour problems; it is a 13-item questionnaire designed for use with children as well as with other people engaging with the children, e.g. parents, teachers, children themselves, and significant others. The BRIC is scored on a 5-point Likert-type scale, with a potential range of scores from 0 to 100. Higher scores indicate more severe behavioural problems. The reported internal consistency for the BRIC is good, with alphas ranging from .80 to .86 for adults and .60 to .70 for children. Test-retest reliability ranges from .71 to .89, but only .50 for children. For the purposes of this study, the Aggression Inventory (also from Fischer & Corcoran 1994) was adapted for use with children, and a test--retest pilot study with five children indicated a reliability alpha of .76 across the measure's four sub-scales and the total AI scores.

One of the purposes of this study was to identify the mechanisms and contexts at the beginning of the work and any changes that take place during the work. The qualitative research strategies applied are based on action-oriented approaches (Rodwell 1998; Drisko 2000; and Taylor & White 2000). The source of all data is the interaction between the client and significant others with the project workers, as well as the cyclical connection between reflection and action. This interaction and reflection is then recorded and categorised within the realist perspective to systematically track mechanisms, contexts and the content of the work. Taylor and White (2000) argue that health and social welfare workers cannot reproduce material situated reality, and what they do is in fact order reality by supporting some versions (e.g. narratives of users and significant others) and burying others. "However, by reflecting on the process of choice and judgement and thinking about ways in which forms may be redesigned to show more of our 'working' we may develop a different and more critical approach to the data contained within them" (ibid., p. 158). In this project, the forms have been designed from the outset using the language of realist evaluation, to enable practitioners to systematically track the outcomes, mechanisms, contexts and the content of interventions.

Reliability of the qualitative data is enhanced by the fact that all sessions with individual clients are co-worked by two workers introducing at least two perspectives, and the team reviews introduce a collective perspective of the team. As Lawson explains, "causal mechanisms that are productive of actual phenomena exist at their own level of being, independently for the most part, of any investigation" (p. 173). However, the perspectives of the practitioner-researcher do influence how they are reflected in the case recordings and data analysis as approximations of the reality. Therefore, one way of dealing with this limitation is the realist effectiveness cycle itself. If, at any one cycle the practitioners have overemphasised certain factors, these limitations will become more apparent in the next cycle in the course of data collection and analysis, and so forth.

The qualitative analysis is then combined with single-subject outcome analysis, to enable an explanation of the effectiveness of Shield's work in each case, particularly in explaining how the causal mechanisms are triggered in the contexts of the clients, and how the causal mechanisms that influence the social problems and behaviour are countered by the alternative causal mechanisms introduced by the Shield project's models of intervention. Therefore, by studying the individual in his/her own causal contexts, the evaluation strategy for each case meets the requirements of intensive research as described by Sayer (2000); but a quantitative element in the form of single-case evaluation is also added to systematically track changes in the outcomes produced in those contexts, as illustrated in the foregoing case example.

Case example: Client K

Client K is a 15-year old male subject of a two-year supervision order for sexual offences with girls both within and outside his family. He has committed a number of serious sexual offences.

The single-case evaluation outcomes for K (table 2) indicate that there were improvements in the repeated scores for Self-concept Scale for Children, Child's Attitude Toward Mother, and Child's Attitude toward Father. There were slight improvements also in Victim Empathy and Adolescent Coping Orientation for Problem Experiences. The only outcome measure to indicate a deterioration was the Aggression Inventory. Table 2 also indicates that the Behaviour Rating Index for Children (BRIC) scores improved from the perspectives of the child, the school, the carers and the parents. In addition to the improvements in the outcome measures, no further allegations have been made against him with regard to inappropriate sexual behaviour. In September (see Box 1) the carers reported more positive behaviours in the children's home, leading to an increase in his freedom and reduction in his supervision. The Shield Project has engaged with K's mother to carry out attachment repair, and she has agreed to have minimal contact with K. In October mother had made no definite arrangements for contact and was very reluctant to visit her son; however, telephone contact was maintained.

Box 1 indicates the qualitative analysis of context and mechanisms, which helps to penetrate underneath the outcomes in table 2. The causal mechanism that appears to shine through is the damaged relationship with both separated parents---the effect of mother's letter on K's motivation and behaviour is an indication of this. This causal mechanism is triggered in the context of the parents' acrimonious separation, and previous history of sexual abuse against both K and his mother. The causal mechanisms that are enabling (in relation to the desired outcomes) include the care placement, school, and the mother's willingness to engage in the work with the Shield project.

The generative causal mechanisms introduced by the Shield project (i.e. the intervention) include exploration of family background and the current situation, discussing the actual offence, processing previous trauma and attachment repair, cognitive approaches to managing behaviour, assessing and developing motivation to change, and trauma outcome process, in separate sessions with K and his mother. Therefore, the improvements in the outcomes as observed in table 2 were achieved with the Shield project's introduction of causal mechanisms that connected with the other enabling causal mechanisms, and all of these causal mechanisms were triggered in the contexts of the parents' acrimonious separation, and history of sexual abuse. These enabling causal mechanisms, as well as the generative mechanisms introduced by Shield, were able to neutralise the disabling causal mechanism, although this battle is by no means over, and the main disabling causal mechanism of K's relationship with his parents still shines through occasionally.

Box 1: Changes in Client K ---REVIEW OF QUALITATIVE DATA

Date of review	Context	Content	Enabling mechanisms	Disabling mechanisms
September 2000	His parents separated acrimoniously and subsequently remarried to different people; K has spent different times living with both parents. At present, K lives in a residential home and has only minimal contact with his natural mother. Both K and his mother have been victims of sexual abuse in the past. K has history of behavioural and academic school problems, but he was not known to social services prior to the offences. He has a history of unmet need when cared for by his family. He suffers from poor social adjustment including isolation from peers, and is himself a victim of sexual abuse and bullying.	Shield provided two female workers holding one-hour sessions twice a week. In just under two months, the content was changed from assessment to treatment work and was provided by one male one female social worker. Two female workers also began work with K's mother. K attended all the sessions, and contributed in all except one.	<ul style="list-style-type: none"> • Appropriate care placement, enjoys it • Mother and partner engaged well with the work. • Settled well, attending school, achieving better performance at school. • K engaging in the work, willing to talk, sense of humour. • Effective multi-agency planning, regular and appropriate liaison. • Mother began work in open and honest manner; some insight into K's needs, able to discuss difficulties. 	<ul style="list-style-type: none"> • Damaged relationship with both parents. • Finds it difficult to express feelings & emotions. • Mother not ready to face the truth about K's possible abuse of siblings, and reluctant to have contact with him. • K anxious about discussing offences or victims, avoids difficult issues. • K using sexually aggressive language with others in the residential home. • Mother had written to K saying did not want any contact with him; K's motivation declined. • Mother wants minimal contact; refused contact despite offer of supported contact. • Father has no contact at all. • Mother and father have limited, acrimonious contact with each other. • Change of workers to male/female, K having to adjust to one new worker
October 2000			<p>All enabling mechanisms continue, other than mother's degree of honesty is not as high as first thought, as she has been unwilling to divulge relevant information which is seen as secrets within the family.</p> <p>Change to include a male worker has been very enabling for K. K is now more in touch with his own feelings and beginning to express them.</p>	Disabling mechanisms all continue; progress is being made on his ability to engage in the work. K is now more in touch with the hurt he feels.

Table 2: Content of sessions with K and his parents; and outcome measures used with K

With Parents	With K	CHILD BRIC	PARENT BRIC	SCHOOL BRIC	CARER BRIC	A-COPE	Self C	C.A.M.	C.A.F	Vic Emp	Agg Inv
	05.03 .	67.5	37.5	45							
3.05	05.05 .						47				55
	09.05 A	47.5	30								
	16.05 A										
18.05 BC	18.05 BCD										
	25.05 .			32.5	30			36	63		
	01.06 CD									15	
06.06 AB	06.06 C										
	09.06 D										
12.06 AB	20.06 B										
	23.06 .			35	32.5						
27.06 .	27.06 .										
	29.06 B1										
	04.07 A1										
	06.07 B1	40					77			17	
	11.07 A1					90					68
	13.07 .										
01.08 .	08.08 A1										
	10.08 .										
22.08 A1	22.08 A1										
	29.08 A1										
06.09 A1	07.09 A1					89	88				
	11.09 B1			25	22.5						
19.09 A1	18.09 B1		15								
	21.09 B1	37.5						25	30	18	68
RESULT		Improved	Improved	Improved	Improved	Improved	Improved	Improved	Improved	Improved	Worsened

Key for Table 2

Outcome Measures

BRIC: Behaviour Rating Index for Children

Self-C: Self-Concept Scale for Children

CAM: Child's Attitude Toward Mother ;

CAF: Child's Attitude Toward Father

Vic emp: Victim Empathy

Agg Inv: Aggression Inventory

A-Cope: Adolescent Coping Orientation for Problem Experiences

Content of sessions (with parents and with K):

Assessment work

A = family background with child

B = current situation with child

C = the actual offence with child

D= changing/understanding the behaviours with child

Treatment work

A1 = processing previous trauma including attachment repair in child

B1= cognitive approaches to managing behaviour with child

C1 = assessing and developing motivation with child

D1= trauma outcome process

In the course of applying the realist effectiveness cycle, both the qualitative and the quantitative analysis was used prospectively to shape and target the content of Shield project's intervention with K. In particular, the identification of the disabling causal mechanisms helped to introduce alternative causal mechanisms which reinforced the other enabling causal mechanisms, in the course of Shield's work. In this way, by integrating realist evaluation strategies, the Shield project workers were able to improve the services provided to both K and his mother.

The systematic tracking of outcomes

The data for each of the measure selected for each client is recorded, as indicated above. The SPSS analysis across the cases enables an account of what works, for whom and in what contexts---the hallmark of realist evaluation. As Bhaskar & Lawson (1998) explain, the effects of sets of causal mechanisms are detectable through the identification of relatively systematic differences in the outcomes of comparable activities in different space-time locations. First, we need to know which outcomes improved, stayed the same or deteriorated. Then, these findings are cross-tabulated with all the rest of the variables with regard to the contexts and mechanisms. For the clients that improved (as well as for those who are in the no change or deteriorated category) with each measure, we need to know what the contexts were, what were the mechanisms and whether these mechanisms were positive or negative, and how these mechanisms changed over time; and what were the components of the content of interventions that were used. At the time of writing, the project evaluation is at an early stage, and therefore there are only eight cases where repeated outcome measures have been used.

Shield Project: discussion

The team as a whole has worked successfully to integrate the realist evaluation approach into their daily practice, to the extent that right from the referral stage, the team examine the contexts and mechanisms for each case, and systematically track the outcomes, as well as changes in the mechanisms and contexts. This data is used to develop the models of intervention with each client as part of a realist effectiveness cycle, and this data is analysed using both intensive and extensive, and both quantitative and qualitative, research strategies. These strategies enable the identification and systematic tracking of demi-regularities in the mechanisms, contexts and outcomes with each client, as well as across all clients, in order to provide an explanation of what works, for whom and in what contexts. SPSS (Foster 1998) is being used to enable cross tabulations of the outcomes, context, mechanisms and content for all cases with repeated outcome measures in the database.

Lawson (1998) defines demi-regularities as "patterns of regularities of sorts, regularities that are recognisable as such despite being something rather less than strict" (p. 162). The reasons why these regularities are not strict are, first, because the environment in which the mechanisms operate may not be homogeneous and there may be a large number of countervailing factors; and, second, the mechanisms themselves are likely to be unstable over time and space. Although demi-regularities are less than strict, and less than universal, the countervailing mechanisms may be such that the primary

mechanisms often dominate. Therefore, at any specific time-space location, there may be systematic and identifiable mechanisms in play which realist evaluation strategies can uncover. However, the central mode of inference is neither inductive nor deductive---it is retroductive. The aim is not to cover a phenomenon under a generalisation, but to identify a factor responsible for it or that helped to produce it. The goal is to "posit a mechanism...which, if it existed and acted in the postulated manner, could account for the phenomena singled out for explanation" (p. 156). This explanatory process also draws heavily on the investigator's beliefs, perspective and experience, e.g. in the way the Shield workers identify the mechanisms and contexts that are to be systematically track over time in their recording system. Nevertheless, as the patterns or demi-regularities are uncovered in the course of the application of the realist effectiveness cycle, the realities are also uncovered to the extent that any relevant mechanism overlooked in the early stages of the work is likely to be discovered over time.

The sub-group analysis presented at this early stage in the evaluation is based on the outcomes achieved in the child's BRIC scores, as an example of how demi-regularities can be identified across clients. Five clients improved when the first and the last (to date) scores were compared---this was the outcome pattern to date. It was also found that school education was an enabling causal mechanism in four of these cases, and that this causal mechanism began to shine through even more in three of these cases. Support from parents was an enabling causal mechanism in three cases, and support from carers in the remaining two. The countervailing or disabling causal mechanisms identified were the client's ability to relate to his/her peers and the relationship between the client's parents, in all five cases. The Shield project workers introduced their interventions that acted as alternative causal mechanisms, harnessing the enabling mechanisms, and neutralising the disabling mechanisms, to produce the outcomes in the child's BRIC. Furthermore, all of these causal mechanisms were triggered in the contexts of the parents' separation, history of domestic abuse, history of physical and sexual abuse against the child, history of school problems as well as a history of poor peer relationships.

The intensive research (as described with the case example of K above) enables the identification of demi-regularities in the particular set of causal factors identified over time with each case. The aggregation of this data and the sub-group analysis enables the identification of demi-regularities across a group of cases. However, this analysis does not remain static, as the mechanisms and contexts of the clients, as well as the alternative causal mechanisms introduced by the project in individual cases, change over time. Also, the extent to which these causal mechanisms are identified and tracked accurately also changes overtime. The analysis presented here, therefore, would change as the numbers of cases are increased, and as the investigation of the clients' realities is developed in this prospective evaluation. In this way, prospective realist evaluation that utilises both quantitative and qualitative methods enables a more thorough investigation of the mechanism-context- outcome configurations over time.

Preliminary findings indicate that all but one of the eight cases with repeated measures to date improved in at least one outcome measure, and four out of the eight cases improved in three or more measures. In terms of cross tabulations, the child BRIC

measure is used as an example of how this methodology will be used in this realist evaluation as the database is extended to include a larger number of cases. The above analysis is not conclusive because of the small numbers, and is included in this paper as an illustration of how the outcomes can be linked with contexts, mechanisms and the content of the interventions in developing evidence-based practice in both assessment and treatment work with children and young people who sexually harm others. Realist evaluation seeks to provide an explanation of an explanation, in an on-going cycle, i.e., why and how certain mechanisms have or have not emerged, and where they have emerged, why they have been "reproduced or transformed in particular ways" (Lawson, p. 162). The analysis presented here is only the first step in this process of realist explanation.

The contribution of realist evaluation for practice

This paper develops realist evaluation as a paradigm for practitioner-evaluators, particularly with the development of a realist effectiveness cycle which can be integrated into the practice of human services and which can penetrate into the realities of practice deeper than the traditionalist view of evidence-based practice. The realist effectiveness cycle is based on the use of both quantitative and qualitative methods, and in the course of its implementation within human service agencies, the repertoire of existing methods are developed and shaped to meet the requirements of this new paradigm.

Study 1 is an example of a largely quantitative studies based on the use of single-case evaluation, or the systematic tracking of outcomes. Sub-group analysis enabled the identification of potential causal mechanisms and contexts through cross tabulations of the outcomes with the demographic characteristics and other circumstances of the service users, as well as some factors associated with the practice of the human service. However, the use of quantitative methods alone across the large numbers of service users included in the study was not sufficient in determining the explanatory retrodution associated with realist evaluations.

In study 2, intensive research (both quantitative and qualitative) is carried out with each service user, to systematically track desired outcomes, enabling mechanisms, disabling mechanisms, the contexts and the content of interventions. The findings from each case are then entered into a database that includes the identification of the common mechanisms and contexts across all service users. First, the realist effectiveness cycle is integrated into the project's practice, and therefore the practitioners utilise realist evaluation to test and develop their models of intervention in an on-going way with each service user. The findings from the systematic tracking of the key dimensions utilising both quantitative and qualitative approaches are used to plan the intervention in each case, in a cycle of development, based on the demi-regularities identified in each case. Second, this data is aggregated across all clients to identify demi-regularities across the wider practice of the project, enabling an account of 'what works, for whom and in what contexts', in the development of replicable models of intervention. This combination of methods enables the ontological penetration and explanatory powers required by a "white box" realist evaluation. This evaluation investigates the causal mechanisms (including the alternative causal mechanisms introduced by the project), associated with the changes in

the outcomes, and the conditions or contexts in which they are triggered (or not), at any time-space location in the practice of the project.

Above all, realism enables you to dig deeper into the embeddedness of social work practice in fluid, unpredictable contexts that is an important and often crucial characteristic of social work practice. In the evaluation of practice illustrated by study 2, the critical realist approach is applied in the development of recording systems, practices and evaluation of effectiveness from the outset. Based on existing knowledge and data accumulation on outcomes, mechanisms and contexts, the programmes are developed as models targeted to achieve the desired outcomes. A multi-method research strategy is applied to test the extent to which these models are analogous with reality, and the data collection and analysis directly contribute to further development of the models as well as their future targeting within a realist effectiveness cycle. In this way, the content of the programmes, their interaction with pre-existing mechanisms and contexts, and their effectiveness in achieving outcomes, are all addressed. The outcomes of the evaluations are to provide data on what works, for whom and in what contexts, along with explanations of why a programme may work with some people and not with others. The findings from the two studies included as examples form the basis for a conclusion that, when compared with other paradigms, critical realism can reach deeper in capturing the shifting sands of social work effectiveness.

References

- Anastas, J. W. (1999) "Research Design for Social Work and the Human Services", second edition, Columbia University Press, New York.
- Archer, M. (1998) "Introduction", in Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (eds.) "Critical Realism: Essential Readings", London: Routledge, pp. 189-205.
- Bhaskar, R. (1997) "A Realist Theory of Science", 2nd. ed., London: Verso.
- Bhaskar, R. (1998) "General introduction", in Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (eds.) "Critical Realism: Essential Readings", London: Routledge, pp. ix-xxiv.
- Bhaskar, R. & Lawson, T. (1998) "Introduction: basic tenets and developments", in Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (eds.) "Critical Realism: Essential Readings", London: Routledge, pp. 3-15.
- Bloom, M., Fischer, J. and Orme, J. (1999), *Evaluating Practice: Guidelines for The Accountable Professional*, 3rd. edition, Allyn and Bacon, Boston.
- Cheetham, J. (1998) 'The evaluation of social work: priorities, problems and possibilities', in Cheetham, J. and Kazi, M. A. F. (eds.), *The Working of Social Work*, London, Jessica Kingsley.
- Cheetham, J., Fuller, R., Mcivor, G. and Petch, A. (1992) *Evaluating Social Work Effectiveness*, Buckingham, Open University Press.
- Drisko, J. W. (2000) "Qualitative Data Analysis: It's Not Just Anything Goes", Paper presented at Society for Social Work and Research Annual Conference, January 30, Charleston, SC, USA.
- Duguid, S. & Pawson, R. (1998) "Education, Change and Transformation: The Prison Experience", in *Evaluation Review*, 22(4), pp. 470-495.
- Fischer, J. and Corcoran, K. (1994), *Measures for Clinical Practice: A Source Book*, vols. 1 and 2, The Free Press, New York.
- Foster, J. L. (1998) "Data Analysis: Using SPSS for Windows", Sage, London.
- Fraser, M., Taylor, M. J., Jackson, R. and O'Jack, J. (1991) 'Social work and science: many ways of knowing?', *Social Work Research and Abstracts*, 27, pp. 5-15.
- Harre, R. (1984) "The Philosophies of Science: an Introductory Survey", second edition, Oxford: Oxford University Press.

- Kazi, M. A. F. (1998) *Single-Case Evaluation by Social Workers*, Aldershot, Ashgate.
- Kazi, M. A. F. (1998a), 'Putting Single-Case Evaluation into Practice', in Cheetham, J. and Kazi, M. A. F. (eds.) *The Working of Social Work*, Jessica Kingsley, London, pp. 187-199.
- Kazi, M.A.F. (1999) "Paradigmatic influences in practice research: a critical assessment", in Potocky-Tripodi, M. & Tripodi, T. (eds.) "New Directions for Social Work Practice Research", Washington DC: NASW Press.
- Kazi, M. A. F. (2000) "Contemporary Perspectives in the Evaluation of Practice". In *British Journal of Social Work*, 30, 755-768.
- Kazi, M.A.F. (2000a) "Evaluation of Social Work Practice in England", *Journal of Social Work Research And Evaluation*, 1 (2), 101-109.
- Kazi, M. A. F., Manby, M. & Buckley, M. (2001) "Evaluation of Family Centres in Kirklees: An Extensive Use of Single-Case Evaluation", *Nationwide Children's Research Centre*, Huddersfield.
- Kazi, M. A. F. & Ward, A. (2001) "Service-wide integration of single-subject designs and qualitative methods: A Realist Evaluation", Paper presented at Society For Social Work & Research Conference, January 2001, Atlanta.
- Lawson, T. (1998) "Economic science without experimentation/Abstraction", in Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (eds.) "Critical Realism: Essential Readings", London: Routledge, pp. 144-185.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000) "Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs", Jossey-Bass, San Francisco.
- Outhwaite, W. (1998) "Realism and social science", in Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (eds.) "Critical Realism: Essential Readings", London: Routledge, pp. 282-296.
- Parton, N. & O'Byrne (2000) *Constructive Social Work: Towards a New Practice*, London: Macmillan.
- Pawson, R. & Tilley, N. (1997a) "An introduction to scientific realist evaluation" in Chelimsky, E., & Shadish, W. R. (eds.) "Evaluation for the 21st Century: A Resource Book", Thousand Oaks: Sage Publications, pp. 405-418.
- Pawson, R. and Tilley, N. (1997b) *Realistic Evaluation*, Thousand Oaks, CA, Sage.
- Phillips, D. C. (1990) "Postpositivistic science: myths and realities", in Guba, E.G. (ed.) "The Paradigm Dialog", Thousand Oaks: Sage, pp. 31-45.

Porter, S. & Ryan, S. (1996) "Breaking the boundaries between nursing and sociology: a critical realist ethnography of the theory-practice gap", in *Journal of Advanced Nursing*, 24, 413-420.

Rodwell (O'Connor), M. (1998) "Social Work Constructivist Research", Garland, New York.

Sayer, A. (2000) "Realism and Social Science", London: Sage.

Scriven, M. (1994) 'The fine line between evaluation and explanation', *Evaluation Practice*, 15(1), pp. 75-7.

Sheldon, B. & Chilvers, R. (2000) "Evidence-based Social Care: A study of prospects and problems", Lyme Regis, Dorset: Russell House Publishing.

Taylor, C. & White, S. (2000) "Practising Reflexivity in Health and Welfare", Buckingham: Open

Thyer, B. A. (1998) 'Promoting evaluation research on social work practice' in Cheetham, J. and Kazi, M. A. F. (eds.), *The Working of Social Work*, London, Jessica Kingsley.

Tolson, D. (1999) "Practice innovation: a methodological maize", in *Journal of Advanced Nursing*, 30(2), 381-390.

Wainwright, S. P. (1997) "A new paradigm for nursing: the potential of realism", in *Journal of Advanced Nursing*, 26, 1262-1271.