

# **PROTOCOLS FOR EVALUATING RESTORATIVE JUSTICE PROGRAMMES**

*Paul McCold, Director of Research at the International Institute for Restorative Practices, Bethlehem, Pennsylvania, USA*

## **Abstract**

This article provides a review and critique of the current research findings about restorative justice. It is suggested that some of the positive findings are not due to programme efficacy, but rather to well-known threats to validity. The effect of case attrition on selection bias is considered in light of the voluntary nature of many restorative justice programs. Standardization of program measures is urged with specific research protocols presented and described. Protocols for measuring participant perceptions are compared. Before scientifically valid statements can be made about best practices, much more rigorous research needs to be conducted. If the results of multiple program evaluations are going to contribute to accumulated understanding of the practice, measures across programs must be standardized. A research agenda is described that would eventually allow for empirically fitting the forum to the fuss and establishing best practice standards across models. Six programme level and six case level measures are proposed as the minimum required for basic program comparisons to be meaningful.

**Key words:** case attrition, selection bias, research protocols, restorative justice, program evaluation, threats to validity

---

Public policy responses to crime should not be based upon the enthusiasm or popularity of programme advocates. The long history of failed criminal justice reform efforts justifies a healthy skepticism. If a justice programme is effective, it should be possible to scientifically measure and convincingly demonstrate these effects. If programme advocates cannot objectively demonstrate the merits of an intervention programme using sound empirical measures, they, too, deserve a large measure of skepticism. Confidence in a given programme's effectiveness becomes appropriate only when positive results are convincingly demonstrated. Confidence in a type or model of practice is justified only after positive results have been replicated in a number similar programmes. There have been nearly 100

restorative justice programme evaluations published as of 2004. Yet, research on restorative justice practice today is a mile wide but only an inch deep (McCold, 2003).

Recent research findings range wildly in their estimates of the beneficial effects of restorative justice programmes, especially regarding claims of reducing offender recidivism<sup>1</sup>. Some researchers conclude that restorative justice is no more effective than court in this regard (Davis, 1982; Roy, 1994; Moore, 1995; Niemeyer & Shichor, 1996; McCold, 1998; Sundell & Vinnerljung, 2004), Others claim to demonstrate moderate reductions in recidivism of 10-15% (Umbreit, 1994; Geudens, 1998; Bonta, Wallace-Capretta & Rooney, 1998; Calhoun, 2000; McGarrell, et al, 2000; Trimboli, 2000; Luke & Lind, 2002; Dölling & Hartmann, 2003; Australian Institute of Criminology, 2004). And some research projects report dramatic reductions of offender recidivism of 30% or more (Chan, 1996; Hsein 1996; Doolan, 1999; Sherman, Strang & Woods, 2000; Wilson & Prinzo, 2001; Rowe, 2002; Chan, 2003). Each of these claims are based upon research protocols with inherent weaknesses, or design flaws that limit the conclusions that can be validly drawn from a given set of outcome results (Campbell & Stanley, 1963).

My own reading of the three dozen studies of reoffending reviewed is that while restorative justice programs do not involve a consistent guarantee of reducing offending, even badly managed restorative justice programs are most unlikely to make reoffending worse (Braithwaite, 2002, p. 61).

Recent efforts to conduct meta-analyses of the findings from restorative justice programme evaluations (Latimer, Dowden, & Muise, 2001; Nugent, Umbreit & William, 2003) are premature, since programmes vary widely in their content, there are too few evaluations that include a valid comparison group, and most programmes have an insufficient number of cases upon which to draw solid conclusions. There are additional significant obstacles preventing any simple aggregation of results from different restorative programme evaluations. The rush to positive judgment presents a danger that conclusions based on unusually cooperative cases may later be shown to be spurious, setting the reform movement up for a perceived failure without a real test of the possibilities.

Demonstrating scientifically valid statistically significant reduction in reoffending by any one-shot criminal justice intervention is rare in the criminal justice literature (Martinson, 1979; Lloyd, Mair & Hough, 1994; Sherman, Gottfredson, et al., 1997; McGuire, 2000). If participation in a restorative process actually does produce consistently lower recidivism rates, it would be unique in that regard and, thereby, destined to become the greatest discovery in criminal justice history with implications well beyond criminal justice (Braithwaite, 1999). Is it possible?

There are really only two primary research questions to be answered about restorative justice. The first is, does it work? The second is who does it work for? Sounds simple enough.

The first research question is about causality. It involves demonstrating a causal relationship between a specific intervention and desired outcomes, primarily changes in reoffending. Before valid conclusions can be drawn about cause and effect, it is necessary to isolate the effects of the programme from the other factors likely to influence future behaviour.

The second research question is about generalisability or external validity. Valid conclusions about the types of cases that benefit most from restorative justice requires that outcomes from a specific intervention model be tested on different types of cases. Available evaluations are of very different programme types used on very different cases making direct comparison meaningless (apples & oranges). This is why results from meta-analysis across currently available evaluation research of restorative justice schemes would also be meaningless (e.g. see Latimer & Kleinknect, 2000, 9).

Specifically, there are a few critical research obstacles that must be overcome prior to drawing valid inferences across programmes. The primary obstacles are 1) dramatically different intervention models, 2) small sample sizes, 3) non-standard measurement protocols, 4) limited client/case characteristics within programmes, and 5) differing amounts and types of selection bias within and between programmes.

Differences in outcomes due to specific characteristics of individual cases has yet to be thoroughly investigated. The effects of known interactions between case specific characteristics (seriousness of offence, age of offender, prior record of offender, victim/offender prior relationship, type of victim, and kind of programme) and both participation rates and outcome measures are certain to exist but the magnitude and direction of these effects remain largely speculative.

Part of the difficulty in detecting possible effects of restorative justice programmes on offender outcomes relates to differential case attrition. Disputes brought to the attention of authorities are not like cases which are not reported to authorities (the so-called 'dark figure of crime'). This introduces a 'reporting bias' into the sample.

Next, those that are reported are screened as appropriate for programme referral or not. Cases selected for referral to a restorative justice programme are not the same as those not referred. For example, offences without an identifiable offender or victim may be excluded. This is the 'system selection bias'.

Finally, among those cases that are referred, not all will agree to participate. Restorative justice programme participation rates range from 11% to above 90% (see McCold, 2003)<sup>2</sup>. Programmes with very low participation rates are likely to contain samples with a higher proportion of cooperative cases than non-participating cases leading to "cherry picking" (selecting only the most worthy cases).

Offenders preferring court are not the same as those who agree to participate; and cases in which the victim also agrees to participate are likely different from cases where they do not. This is the 'self-selection bias'.

It is worth noting here that the existing criminal justice system contains these same case screening and sample bias problems as restorative justice programmes. Crimes are unfounded by the police, withdrawn by prosecutors, dismissed and suspended by judges. Plus, many offenders simply do not appear in court in spite of the possibility of an arrest warrant being issued:

[T]here are too many cases there to begin with that end up in dismissals and fines that are impossible to collect and they should be diverted in the very first place. ... If you can somehow relay ... the cost of processing cases that end up going nowhere to no satisfactory result to any of the participants in the system, then you can show that mediation is not only worthwhile but absolutely necessary (MacWillie 1981:34-5).

Recidivism is not independent of the decisions to report an offence, refer a case, to participate in a programme or to comply with a court order. Because of differential case attrition, results achieved in cases participating in restorative justice programmes cannot be presumed to work as well with all referred cases, all known cases or all cases. Each represents a distinct sampling frame.

McCold and Wachtel (1998) found that offenders who were willing but unable to participate in a restorative justice conference were less likely to reoffend than those not wanting to participate. If this effect holds true for other voluntary programmes, as seems likely, then any sample of participating offenders should always compare favorably since it excludes the uncooperative offenders, while the comparison groups contain a mix of cooperative and uncooperative cases. The difference in reoffending is the result of the decision to participate rather than the effects of the programme itself. While this is encouraging for justifying case diversion through restorative programmes, it creates a difficulty in interpreting findings from research assessments that use match comparison groups unless those not participating are also included.

The factors that determine offender willingness to participate are largely unknown. McCold & Wachtel (1998) found that girls were more likely than boys to agree to participate in a conference, but the decision was not related to race, age, or seriousness or type of offence. Since recidivism is related to offender self-selection, measures of offender characteristics likely to affect this decision are needed. For example, it may be that personality characteristics like level of self-control (Gottfredson & Hirschi, 1990; Hirschi, 2000) determines whether an offender perceives a given process as fair (Piquero, Gomez-Smith & Langton, 2004) and, thereby, tip the balance between feeling shame and becoming defiant (Sherman, 1993). That is, high self-control offenders may be more willing to participate in restorative justice and less likely to reoffend than offenders with

low self-control (Tyler, 1990; Kinsey, 1992; Makkai & Braithwaite, 1994). Offenders with high self-control are more likely to be remorseful and to take responsibility for their behaviours than are offenders with low self control (Gottfredson & Hirschi, 1990, 93), regardless of whether or not they participate in a restorative process.

The standard way to control for selection bias is through random assignment of cases to programmes after the decision has been made to participate. Random assignment does not remove the selection biases but assures that it is equally distributed among the assigned groups. The stage at which the cases are randomised becomes the sampling frame to which the results can be generalised. If a high percentage of cases assigned to a programme actually participate, then valid conclusions can be drawn about the effects of this programme on this sample (i.e., cooperative offenders). But what about a different type of case (i.e. uncooperative offenders)?

This has not been a problem for research on juvenile cases in New Zealand, which mandates a conference for nearly all serious juvenile offenders. The problem here is the lack of a comparison group. In a multivariate analysis of participating cases, Maxwell and Morris (2001) found that reoffending was related to offender early childhood experiences, the impact of the conference, and life events afterwards. They conclude that having a conference is, by itself, unlikely to be effective. The critical variables for a successful conference are 1) whether or not the conference was a memorable event, 2) that it evoked remorse and 3) this led to the young person meeting the victim, apologizing and attempting to make amends. These factors remained significant after controlling for the other factors (Maxwell & Morris 2001:253).

Daly and Hayes (2003) similarly conducted a multivariate analysis of conferencing data from South Australia Juvenile Justice programmes and found that prior offending, sex, race and social marginality were highly predictive of post-conference offending. After controlling for these, re-offending was still found to be lower for offenders who expressed remorse and whose outcome decisions were consensual.

The difficulty with these single sample multivariate analysis is that they may be observing the effects of offender self-selection and mistake the direction of causality. Offenders who are remorseful are more likely to meet their victim and to want to make amends. Offenders who are not remorseful will resist meeting their victim, acknowledging their offences, or cooperating toward solutions. Whether or not they attend a conference, it is more parsimonious to assume that the offenders' level of remorse determines both their cooperativeness and likelihood of reoffending. Thus, the characteristics of conferences listed above may be spuriously related to reoffending since they are both caused by the offender's attitude toward their behaviour. This self-selection effect is exacerbated in those restorative programmes that require offenders to accept responsibility as a precondition for participation.

Beyond documenting a self-selection effect on recidivism, few assessments have explored the relationships between case attrition and outcomes within or between programmes or case types. Evidence that certain types of cases produce agreements which are “easy to get but hard to keep” remains mixed. While conferencing programmes seem to produce higher outcomes on average than mediation (McCold & Wachtel, 2002; but also see Umbreit, Coates & Vos, 2000), without cross programme multivariate assessments of programme, dispute and disputant characteristics, there is insufficient empirical basis for establishing “best practices”. Until results from controlled cross-programme trials complete with cost-benefit analyses are available, no empirical criteria exist for determining how to best “fit the forum to the fuss”.

Thus far, I have avoided the obvious problem related to the first research question of Does it work?, which is, What constitutes working? How to quantify success? Comparisons of any new approach, such as restorative justice, must be measured against existing practice. While this may seem obvious, there is a tendency to compare new programmes to perfection and criticise them when they fail to reach the ideal. To succeed, restorative justice does not need to be perfect. To be preferred, a programme need only demonstrate superiority, on average, to traditional adjudicatory approaches. Thus current justice systems set the benchmark against which restorative programmes should be tested.

Evaluating a new paradigm by the sole criteria of the old is inappropriate. Restorative justice offers advantages for crime victims, communities, and offenders families, even if the practice is eventually shown to have no direct affect on offender recidivism rates. Offenders are expected to be held accountable for the consequences of their misbehaviour, but as a way to begin to address the offender's need to learn responsible behaviour. Holding offenders accountable in a reintegrative manner is expected to affect their future behaviour, but changing that behaviour is not the primary purpose of restoration (McCold, 1996). The goals of restorative justice are to meet the real needs of victims, offenders, and their communities created by the criminal act (McCold, 2004).

## **What Should be Measured?**

Latimer and Kleinknect (2000, p. 9) suggest research be conducted using two distinct frames, participant level and system level. Measures at the participant level include satisfaction, recidivism, and sense of security for victims, community and offenders. Research at the system level includes cost-benefit analysis, crime statistics data collection and effects on police, attorneys, courts and corrections:

...the central goal of the criminal justice system should be to reconcile victims, offenders and their communities while repairing the harm caused by the criminal behaviour. That is not to say that public safety is not paramount. Rather, it is the method of achieving public safety that is under debate (Latimer & Kleinknect, 2000, 6).

The most direct way to measure the restorativeness of a process is to ask the stakeholders if their needs were met. Most often, restorativeness has been measured (i.e. operationalised) in the research by surveying participant perceptions. If victims reparative needs are met, then they should be satisfied, if not they will not be satisfied. If offenders responsibility needs are met, they should feel the process was fair. If the families' needs are met, they should also express satisfaction with the process and think it was fair. McCold and Wachtel (2002) demonstrated that fully restorative programmes tended to have the highest and most balanced results from both victim and offenders.

No studies have evaluated the effect of restorative justice on the nature and strength of social bonding within localities receiving intensive restorative services over an extended period. The aggregate reduction in the crime rate reported from Sparwood, British Columbia, Canada (Bouwman & Purdy, 1997) have not been replicated elsewhere. Thus, the aggregate crime reduction potential of restorative justice remains unexamined.

There are a number of important moderating (or confounding) variables that can affect the outcomes of a programme, including age of the offender and seriousness of the offence, the models of intervention, the entry point of programme, whether the intervention is a justice or social service, the context of offence, the background, age and gender of participants, prior relationships, facilitator level of training, and whether facilitators are volunteer or paid.

I would urge the research tasks be seen at two levels of aggregation, the individual level and the programme level. Only after a restorative programme has become established enough to process a large portion of total cases in a given jurisdiction are measures of system level variables likely to show changes. Also, programmes capable of processing large numbers of cases are likely to have different characteristics to small selective programmes, so one cannot infer that outcomes from small pilot programmes can be reproduced on a grander scale.

For research results to be comparable across programmes, details on the nature of the intervention, the characteristics of cases, as well as the immediate and longer term outcomes are needed in a way that is consistently measured. What variables should be included in a restorative justice programme evaluation and how shall these variables be measured?

Presented below is a list of my recommendations for variables to measure within a moderate research budget. Measurement of these variables would provide sufficient information to address the majority of research difficulties discussed above.

A. independent variables:

- type of intervention
- purpose of intervention
- point of intervention
- conditions for programme eligibility
- type of facilitator

B. intervening variables:

*participants*

- age, gender and race of offender and victim
- prior relationship between victim and offender
- offender risk of reoffending
- offender level of self-control
- family support for victim and offender
- existence of ongoing underlying problems

*process*

- number of participants present
- voluntariness of participation
- adequacy of participation

*programme*

- volume of cases and the participation rate
- cost/time per case

C. dependent variables:

- agreement rates
- compliance with agreement rates
- satisfaction rates
- fairness rates
- victim fear of offender
- victim and offender sense of support
- offender rearrest rates at 12 & 24 months

variables disaggregated by:

- offender participates, fails to participate, or declines to participate
- victim present, indirect participation, or absent
- individual or corporate victim
- juvenile or adult offender
- crimes against person, property, or public order

## How Shall These be Measured?

There is a tendency for researchers to over-complicate the task of operationalising the independent, intervening and outcome variables. The easiest way to measure someone's opinion is to ask them. If we want to know if victims feel satisfied, ask them. For comparison purposes, it matters less whether we use a yes/no question, a five-point semantic differential, or a zero to 100 scale, as long as a percent positive response can be computed.

Perceptual measures should be gathered on all victims, offenders, and their families, for those participating and those declining to participate, and any professionals or volunteers involved, including the facilitator. Victim view of offender and victim and offender perceptions of social supports should be measured after a period of time has past after the restorative intervention (e.g. 6 months), and again after two years to test the stability of the effect. Rearrest rates should be gathered on participating and referred non-participating offenders and a suitable comparison group. Recidivism rates should be reported at 1, 2, and 3 years, where possible. Re-offending is best measured at the earliest point of detection (normally at arrest). Recidivism measured at later points in the system suffer increasingly from the same selection biases discussed above.

For most purposes, limiting answers to yes or no is adequate. Reporting programme results only as a mean (e.g. 8.5 on a 10 point scale) does not allow for comparison. The percent responding positive could easily be computed from the raw data and reported as well. It would be simpler to use a forced-choice response (i.e. without a neutral response). No one is truly neutral and a neutral response cannot count as a positive in computing percent positive. This may turn out to not matter much, since well run restorative programmes have satisfaction and fairness ratings approaching 100% positive. Most variation in response is limited to the agree and strongly agree categories. Not too much should be made of these differences, since marking the extremes on a scale may tell more about how people mark answers than about their perceptions of the programme.

Participant perceptions should generally be measured about the entire process to be comparable between programmes. Evaluations may also wish to include additional questions to distinguish between process and outcome measures. Asking victims "Overall, are you satisfied with the way your case was handled?" is better than asking "Are you satisfied with the outcome?" The former more likely taps into perceptions of the programme, while the latter measures response to the specific details of the agreement.

Measurements of the independent variables are relatively straight forward. The type of intervention should be specified as circle, conference, FGDM, mediation, board/panel, or other (empathy training, victim only service, etc.). Any variation in practice or deviations from accepted protocols of practice of a particular model should be noted. Whether the intervention is diversionary or not should be noted. The number of cases from different intake sources or the points of intervention should be also be reported. These include the

number of walk-in, pre-arrest, pre-prosecution, pre-sentence, in-programme, and post-release applications of the intervention. Any specific conditions for programme eligibility and grounds for excluding cases should be noted. General qualifications, training and type of facilitator (e.g. volunteer, social worker, police, other professional) should be noted.

The intervening variables should be measured as either numerical counts or simple categories. Age and gender are self explanatory. Race (or ethnicity) need only indicate if victim or offender are members of a distinct minority. Distinguishing between more than two or three racial groups would not normally be necessary. Prior relationships between victim and offender should be categorised appropriately (e.g. strangers, occasional acquaintances, regular acquaintances, friends, or intimates/cohabitants). More elaborate schemes may be appropriate, but should be reported in such a way the categories can be collapsed into these simpler ones to allow for determining if violations of relationships of trust were involved (Fatic, 1995). Categories need to be mutually exclusive and exhaustive.

To measure relationships with families and communities of care, at a minimum, victims and offenders might be simply asked to rate their relationship with their family and friends before and after the intervention. If restorative interventions work in the way theory predicts, the strengthening of social bonds is a key mechanism to any longer term changes. It is important to evaluate how the level of social support is affected by the intervention compared to what would have happened without the intervention. Of course, much more extensive evaluation of family systems would also be fruitful, resources permitting.

Research suggests that cases with protracted underlying problems, such as a history of family violence, drug or alcohol use, or mental health issues often lead to agreements which call for specific behavioural changes such as attending counselling (Roehl & Cook, 1989). These tend to be the agreements that are easy to make, but hard to keep. Expectations for these types of cases and the appropriate type of intervention should be different than among cases without these complications.

The total volume of cases handled by a programme is important to assess whether similar outcomes might be expected of a similar programme handling significantly larger numbers of cases. The number of referrals and the number participating in both direct and indirect processes should be reported. Finally, the number of participants at the restorative intervention should be recorded separately by their role (victim, offender, offender supporters, victim supporters, community member, agency representative, etc.), including whether the process was co-facilitated or not. Related to this, programmes should attempt to measure the staff and facilitator resources (time per case) invested in each case, listing separately the preparation for the intervention, the intervention process itself (and length of private family time, if any), and any staff time for case follow-up. These measures should be compiled both on cases completed and terminated early, with the reason for termination noted.

Formal risk assessment for offenders is useful to predict who is likely to reoffend, not who would be a good candidate for a restorative intervention. Risk of reoffending should not, by itself, be a barrier to programme participation, not if we are to learn whether restorative approaches reduce these pre-existing risks. Existing level of offender risk is a useful variable to set expectations about recidivism rates and how an intervention might affect those rates. Knowing how many times an offender has been arrested, age at first arrested, race and gender accounts for most of the variation in reoffending rates, and are sufficient measures to control for the intervening effect of history on recidivism. Tests of recidivism between treatment and comparison groups should always control for prior risk. With sufficient research budgets, offender level of self-control should be measured much more thoroughly using available scales designed for this purpose (Grasmick et al., 1993; Pratt & Cullen, 2000; Weibe, 2003).

Knowledge of whether participants felt pressure to participate would allow for interpreting the effect of coercion on the outcomes. Victims and offenders should be asked if participating was their own choice, and if they felt pressure, asked the source. Responses might be: Yes, No, and Not really, and, if negative, was the source of pressure from officials or from their family? Lastly the perceived adequacy of participation should be assessed for victim, offender, families, professionals, and others. Asked, "Were you able to say what you needed?" with "yes" or "not really" responses would suffice.

The percentage of cases resulting in an agreement (whether written or not), the general nature of the agreements and the number of these agreements complied with fully, partly, not complied with, or pending completion should be reported annually. Metrics for measuring participant satisfaction and fairness were discussed above. There should be a preference for simplicity, or at least the possibility of simplification, in the way questions are asked (i.e. percent positive computable). There are, of course, reasons for much more detailed information to be collected on changes experienced by participants, but particularly for victims. This is especially true in serious cases where victims have been traumatised by the offence. Recent research suggest the benefits of restorative justice to crime victims may be the most compelling justification for the approach (Strang, 2001).

An important research consideration is when to measure outcomes. For example, the victim's satisfaction with the way their case was handled will likely change if the agreement is not kept. Yet waiting to survey participants to allow time for agreements to be fulfilled will result in a lower return rate, adding a reporting bias into the results. Measuring participants immediately following an intervention ensures a high return rate, but may result in inflated ratings due to an afterglow effect reported by some restorative interventions. What to do?

Let us consider this tradeoff more closely. The threats to the validity of participant surveys that are distributed immediately following the intervention may be inflated due to:

1. a false temporary satisfaction that fades with time after reflection, a so-called bubble effect,
2. actual attitudes of participants change (e.g. due to the offender not fulfilling the agreement), and
3. participants give answers they think researchers want.

In research, there are always trade-offs between the validity of the results and the cost and time in collection of the data. Threats to validity need to be considered within a cost-benefit perspective. Researchers considering whether to survey participants face some simple choices.

Choice 1: Ideally participants should be interviewed in person by a researcher unconnected with programme operations with a guarantee of anonymity, followed by repeated attempts to complete interviews. This is the most expensive and time consuming survey method, but is likely to produce the most scientifically valid results. The Reintegrative Shaming Experiment (Sherman, Braithwaite, et al., 1997) interviewed participants six months and twelve months following the conference (and court). Thus their response rate was very high, limited only by the inability to track down a few people. This type of research may be more expensive than the programme it is evaluating.

Choice 2: In the Bethlehem Experiment (McCold & Wachtel, 1998), postal surveys were sent to participants six months following the conference. This got the usual response rate for postal survey research (40-50%). This was followed-up phone interviews for those not responding by post, and in-person interviews were conducted for those who could be located who had no phones. This additional time and expense, produced a response rate of 75%. There was no difference in the response between those answering by post, by phone, or by personal interview.

Choice 3: Another choice for surveying participants is to rely only on postal surveys without follow-up, knowing the potential for nonresponse bias increases as response rates decrease. This is a relatively inexpensive but common survey methodology. However, with participant satisfaction and fairness running above 90%, it would only take a few dissatisfied non-respondents to make the difference between, say 99% and 94% (depending upon sample size).

Choice 4: Another option is to give questionnaires to participants in self-addressed stamped envelopes when they leave the conference. Participants could be asked to return the survey after they have had time to reflect on the conference. This would likely produce a response rate better than posted surveys because all participants have at least been provided with the survey (eliminating the problem of tracking down subjects later). This would also minimise the concern about response reactivity, but would not account for later failure to comply with the agreement and still may have disgruntled participants not responding.

Choice 5: Lastly, participants could be asked to fill out short questionnaires just prior to their leaving the conference setting. This would provide the highest response rate and therefore the least nonresponse bias. However, this would fail to measure attitudes changing with time for reflection or because offenders fail to meet expectations.

All five of these options are acceptable research practices. They are listed here in descending order of cost.

Like RISE (Strang, 2002), the Bethlehem Experiment (McCold & Wachtel, 1998) also found that victims were much less likely to express satisfaction with the conference when offenders violate the terms of agreements (6% in our case). This is NOT the bubble effect, but an actual change in perceptions. However, since 90+ percent of conferenced offenders comply fully with agreements, the effect of noncompliance on participants perceptions can be taken into account, simply by re-surveying the participants in these few cases.

The Queensland conferencing pilot (Palk, Hayes, & Prenzler, 1998) surveyed 456 victims, offenders, and parents immediately upon conclusion of the conference, and 204 again 2-4 months later. The response rate fell from 100% to 69% (204 of the 294 scheduled for follow-up). Yet, the differences in perceptions were not significant. Asked about satisfaction with the conference agreements, offenders went from 99.1% to 100%, victims from 96.7% to 92.1%, parents from 98.6% to 98.8%. Asked about fairness, the offenders went from 98.2% to 100%, victims 98.2% to 93.7%, parents unchanged at 100%. Finally, would they recommend conferencing to others, offenders went from 98.2% to 100%, victims from 98.8% to 96.8%, parents from 99.3% to 98.8%. Thus, the bubble effect was not evident.

Ideally, measures of victim and offender sense of their degree of social support would be measured prior to the intervention, immediately afterwards, and again perhaps six months later. This would allow for a test of the effect of the intervention on the strengthening or weakening of social bonds as predicted by theory (Braithwaite, 1989).

Full compliance with these protocols is possible for research evaluations with moderately large research budgets. Of course, not all programmes can afford to conduct extensive evaluations. Some programmes may opt to measure only the bare minimum. What are the most important variables for making the results comparable to other programmes? Even if programmes choose not to survey participants or track recidivism, at a minimum all programmes should report these data per year at the programme level:

1. the type of intervention (e.g. mediation/circle/conference with/without private time)
2. number of cases referred
3. sources of referral
4. number of cases participating
5. number of agreements reached
6. number of agreements completed

All programmes should collect these data at the case level and also report the aggregate results on an annual basis.

1. the type/nature of case (e.g. adult/juv, diversionary or not, theft/assault/other)
2. facilitator time in preparing conference
3. length of conference
4. number of participants
5. nature of agreement
6. the reason for cases referred but not conferenced, (offender/victim refused, etc.)

For very little additional expense, simple one-page questionnaires for participants to fill out at the end of the intervention rating their satisfaction and sense of fairness would greatly add to the findings and provide a minimal metric of success from a restorative perspective.

## Conclusions

Different restorative justice models do not function equally well with all types of cases. Before we can begin better “fitting the forum to the fuss” or draw conclusions about which of the variations in the processing of cases constitute “best practices” for any type of case, programme assessments involving random assignment of different case types to different resolution processes are needed (Roehl & Cook 1989:37). “What works best for whom?” is an empirical question. Scientifically fitting the forum to the fuss requires a randomised factorial design with a sufficiently large number of cases and an adequate research budget. It is likely that such an experiment on restorative justice models will eventually be conducted by someone, somewhere in the world. In the meantime, the practice of restorative justice remains more an art form than a science.

If research on restorative justice practice is to evolve, findings need to be comparable across programmes. Assessments need to report conditions for programme eligibility, the number of cases referred, the number of cases participating, reasons for nonparticipation, number and nature of agreements reached, and the rate of compliance with agreements. Results need to be disaggregated by offence type and disputant relationship. Programmes providing indirect conciliation and conferences without victim presence should report these data separately. Referred but nonparticipating cases need to be included in follow-up surveys and incorporated into the analyses of programme outcomes. Participant satisfaction/fairness findings should be gathered in a consistent way and reported in a manner that allows for computing the percent positive response.

And, of course, there should be monitoring of programme integrity: such that they adhere to their appointed objectives and selected methods of intervention. Badly-designed, poorly implemented services will emerge as ineffective regardless of the restorative model or the criminal justice setting (McGuire, 2000). Without concerted governmental guidance, programme assessments are likely to continue to reconfirm already well established findings while contributing little to the cumulative understanding of the practice or theory of restorative justice.

There are a few basic lessons that seem reasonable to draw from the collection of all the research on all the different restorative justice programmes that does directly address best practices issues. From my favorite restorative justice scholar, John Braithwaite:

Western ADR (alternative dispute resolution) models invented in the 1970s were going nowhere slowly during the 80s until they learnt some crucial lessons from indigenous justice in the 1990s. In the criminal justice context, there were three most crucial lessons. The first was that dyadic mediation between a victim and an offender is an impoverished formula for restorative justice compared with bringing into a circle a multiplicity of people who are affected in different ways, but particularly people who love and want to support those affected as victims and offenders. ...The second lesson ... that it is better to put the problem rather than the person in the centre of the circle. ...The third crucial insight was that material reparation was less important than symbolic or emotional reparation (Braithwaite, 1999, 2-3).

In spite of the research obstacles, results to date provide strong empirical support for a few additional generalisations, at least until future empirical results find the contrary. Disputing parties typically hold positive views of restorative justice programmes; they feel satisfied with the process and would return if a dispute arose in the future. Studies involving different settings and types of disputes found disputants perceived the outcomes of restorative justice to be significantly fairer than those of court proceedings.



There is no significant public opposition to restorative justice. There is a high degree of support among victims of crime and the public for offender reparation and for victims to have an opportunity to meet with their offender. While participation rates vary widely from programme to programme, most victims and offenders will choose to participate given the opportunity. Offenders are somewhat less likely to participate than victims. Participation rates differ for type of offences, age of offender, type of victim, and the relationship between victim and offender.

Many mediation programmes offer the option of indirect mediation (conciliation). Victims rate direct mediation higher than indirect mediation or court, and reported being less fearful when they meet their offender face-to-face (Umbreit & Roberts, 1994).

At least one conferencing programme, in New Zealand, requires participation of qualifying juvenile offenders (Maxwell & Morris, 1994). This programme and some other conferencing programmes proceed with or without the attendance of the victim. Victim participation and satisfaction rating for New Zealand's youth justice family group conferencing was lower than for most restorative programmes (McCold, 2003).

Where victim and offender participate in restorative programmes, the rates of agreement and compliance with that agreement are very high. Victims and offenders with ongoing relationships produce agreements that are easy to get, but hard to keep. There is no consistent relationship between a programme's participation rate and either the agreement or compliance rates.

There is no intrinsic limitation to the type of dispute or disputants for which restorative justice can bring a reparative response and no empirical limitation reported in the evaluation research. Mediation and conferencing have reported successful resolutions in violent and property cases, adult felony and first time juvenile cases, and between strangers or among family members. For everything from consumer complaints to domestic violence, programme evaluations have documented the positive outcomes of restorative justice.

Justice does not need to be a trade-off between victims and offenders. Both victims and offenders rate restorative justice as more fair and satisfying than court. This is especially true for victims and models that directly involve communities of care. Several recent restorative justice programmes report fairness and satisfaction ratings from both offenders and victims above 95 percent.

Reoffending rates for offenders is no higher for restorative justice than it is for court adjudication. The effects of the programme on reoffending depends upon crime type and is related to participation rates. While there appears to be a strong self selection effect for the voluntary programmes, reoffending following restorative justice processing seems to be reduced more among offences against the persons, than property offences or victimless offences.

Despite the limitations of the existing research, the overall results from the empirical findings has been convincing that restorative justice programmes do consistently perform as well or better than traditional prosecution approaches across different countries, clients, and programme types. Given these findings, restorative justice can be expected to continue to gain popular and political support. The future development of restorative justice should be accompanied and guided by carefully controlled scientific assessments using standardised measures across a variety of settings and practices. Only then can questions about what works best with whom be addressed with anything more than opinion.

## End Notes

1. Research should demonstrate the other tangible benefits produced by restorative justice interventions as well, of course. "Restorative justice mostly works well in granting justice, closure, restoration of dignity, transcendence of shame, and healing for victims. It also shows great promise as a strategy of crime reduction" (Braithwaite, 2002, p. 69).
2. The practical limits of a voluntary approach remain undetermined. For example, how does compelling offender participation effect victim participation rates, participant perceptions, agreement compliance, and offender recidivism? The finding of very poor victim perceptions of New Zealand's mandatory FGCs may be partly explained by the lack of offender voluntariness (i.e. self-selection does not take place, so unrepentant offenders are not able to opt out).

## References

- Australian Institute of Criminology. (2004, February). Restorative justice as a crime prevention measure. *AIC Crime Reduction Matters*, 20. Australian Institute of Criminology.
- Bonta, J., Wallace-Capretta, S, Rooney, J., & Mcanoy, K. (2002). An Outcome Evaluation of a restorative justice alternative to incarceration. *Contemporary Justice Review*, 5(4), 319-338.
- Bouwman, J., & Purdy, G. (1997, April). Sparwood Youth Assistance Program. Paper presented to Dawn or Dusk in Sentencing, CIAJ National Conference, Montreal. <http://www.sparwood.bc.ca/syap.htm>
- Braithwaite, J. (1989). *Crime, shame and reintegration*. New York, Cambridge University Press.
- Braithwaite, J. (1999). Assessing an immodest theory and a pessimistic theory. In M. Tonry (Ed.), *Crime and justice, Vol. 25: An annual review of research*. Chicago: University of Chicago Press.
- Braithwaite, J. (2002). *Restorative justice and responsive regulation*. New York: Oxford University Press.
- Calhoun, A. (2000). Calgary Community Conferencing. Activities and outputs: January 1998-July 2000. Calgary, Alberta: Calgary Community Conferencing.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
- Chan, W.-C. (1996). *Family conferences in the juvenile justice process: Survey on the impact of family conferencing on juvenile offenders and their families*. Singapore: Subordinate Courts Statistics and Planning Unit Research Bulletin.
- Chan, W.-C. (2003). Victim-offender mediation, Making amends and restorative justice in Singapore. In T. Ota, (Ed.), *Victims and criminal justice: Asian perspective* (pp.227-260). Tokyo: Hogaku-Kenkyu-Kai, Keio University.
- Daly, K., & Hayes, H. (2002). Youth justice conferencing and re-offending. Revised paper presented at the Australian and New Zealand Society of Criminology 16th annual meeting, Melbourne.
- Davis, R C. (1982). Mediation: The Brooklyn experiment. In: R. Tomasic and M. Feeley (eds.), *Neighborhood justice: Assessment of an emerging idea* (pp. 154-170). New York: Longman.

- Döilling, D., & Hartmann, A. (2003). Re-offending after victim-offender mediation in juvenile court proceedings. In E. Weitekamp & H.-J. Kerner, (Eds.), *Restorative justice in context: International practice and directions* (pp.208-228). Devon, UK: Willan Publishing.
- Doolan, M. (1999). The family group conference – 10 years on. Paper presented at the Building Strong Partnerships for Restorative Practices. Burlington, Vermont.
- Fatic, A. (1995). *Punishment and restorative crime-handling: A social theory of trust*. Aldershot, UK: Avebury Ashgate Publishing.
- Geudens, H. (1998). The recidivism rate of community service as a restitutive judicial sanction in comparison with the traditional juvenile justice measures.. In L. Walgrave (Ed.), *Restorative justice for juveniles: Potentialities, risks and problems for research* (pp. 335-350). Leuven, Belgium: Leuven University.
- Gottfredson, M., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Grasmick, H., Title, C, Bursik, R, & Arneklev, B. (1993) Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of Research in Crime and Delinquency*, 30, 5-29.
- Hirschi, T. (2000). In defense of self-control. *Theoretical Criminology*, 4, 55-70.
- Hsien, L.L. (1996). Family conferencing good for young delinquents: Report. *Straits Times*, March 6.
- Kinsey, K. A. (1992) Deterrence and alienation effects of IRS enforcement: An analysis of survey data. In J. Slemrod (ed.), *Why people pay taxes*. Ann Arbor, MI: University of Michigan Press.
- Latimer, J., Dowden, C., & Muise, D. (2001). *The effectiveness of restorative justice practices: A meta-analysis*. Research and Statistics Division Methodological Series. Department of Justice, Ottawa, Canada.
- Latimer, J, & Kleinknecht, S. (2000, January). The effects of restorative justice programming: A review of the empirical. Ottawa: Department of Justice Canada, Research and Statistics Division.
- Lloyd, C., Mair, G. and Hough, M. (1994) Explaining reconviction rates: A critical analysis. Home Office Research Study No.136. London: HMSO.
- Luke, G., & Lind, B. (2002). Reducing juvenile crime: Conferencing versus court. *Crime and Justice Bulletin*, 69, 1–20.
- MacWillie, J. (1981). *The price of justice: The cost of arresting and prosecuting three robbery cases in Manhattan*. New York: New York City Criminal Justice Coordinating Council.
- Makkai, T., & Braithwaite, J. (1994). The dialectics of corporate deterrence. *Journal of Research in Crime and Delinquency*, 31, 347-373.
- Martinson, R. (1979) New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review*, 7, 243-258.
- Maxwell, G., & Morris, A. (1994). The New Zealand model of family group conferences. In C. Alder and J. Wundersitz (eds.), *Family conferencing and juvenile justice: The way forward or misplaced optimism?* (pp. 15-44). Canberra, AUS: Australian Institute of Criminology.
- Maxwell, G., & Morris, A. (2001). Restorative justice and reoffending. In H. Strang and J. Braithwaite (Eds.), *Restorative justice: Philosophy and practice* (pp. 93-103). Burlington, VT: Ashgate Publishing Company.
- McCold, P. (1996). Restorative justice and the role of community. In B. Galaway & J. Hudson (Eds.) *Restorative justice: International perspectives* (pp. 85-102). Moneys, NY: Criminal Justice Press.
- McCold, P. (1998, March). Restorative policing experiment: Recidivism results. Paper presented at the Academy of Criminal Justice Sciences, Albuquerque. [NCJ 177565]
- McCold, P., & Wachtel, B. (1998). *The Bethlehem Pennsylvania police family group conferencing project*. Pipersville, PA: Community Service Foundation.
- McCold, P. & Wachtel, T. (2002). Restorative justice theory validation. In E. Weitekamp and H.-J. Kerner (Eds.), *Restorative justice: Theoretical foundations* (pp. 110-142). Devon, UK: Willan Publishing.
- McCold, P. (2003). An experiment in police-based restorative justice: The Bethlehem (PA) Project. *Police Practice and Research*, 4(4), 371-382.
- McCold, P. (2003). A survey of assessment research on mediation and conferencing. In L. Walgrave (Ed.), *Repositioning restorative justice* (pp. 67-120). Devon, UK: Willan Publishing.
- McCold, P. (2004). The role of community in restorative justice practice and theory. In H. Zehr & B. Toews (Eds.), *Critical issues in restorative justice*. Monsey NY: Criminal Justice Press.
- McGarrell, E., & Olivares, K., Crawford, K., & Kroovand, N. (2000). *Returning justice to the community: The Indianapolis juvenile restorative justice experiment*. Indianapolis, IN: Hudson Institute.
- McGuire, J. (2000-July). What works in reducing criminality. Paper presented at the Reducing Criminality: Partnerships and Best Practice conference. Canberra, Australia: Australian Institute of Criminology.
- Moore, D B. (1995). A new approach to juvenile justice: An evaluation of family conferencing in Wagga Wagga. A Report to the Criminology Research Council. Wagga Wagga, NSW: Centre for Rural Social Research, Charles Sturt University-Riverina.
- Niemeyer, Mike, & Shichor, David. (1996). A preliminary study of a large victim/offender reconciliation program. *Federal Probation*, 60(3):30-34.
- Nugent, William R., Umbreit, Mark S., & Williams, Mona. (2003). Participation in victim-offender mediation and the prevalence and severity of subsequent delinquent behavior: A meta-analysis. *Utah Law Review*, 2003(1), 137-166.
- Palk, G., Hayes, H., & Prenzler, T. (1998). Restorative justice and community conferencing: Summary of findings from a pilot study. *Current Issues in Criminal Justice*, 10(2):138-155.
- Piquero, A., Gomez-Smith, Z., & Langton, L. (2004). Discerning unfairness where others may not: Low self-control and unfair sanction perceptions. *Criminology*, 42(3), 699-733.
- Pratt, T., & Cullen, F. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology*, 38, 931-964.
- Roehl, J., & Cook, R. F. (1982). Neighborhood Justice Centers field test. In R. Tomasic & M. Feeley (Eds.), *Neighborhood justice: Assessment of an emerging idea* (pp. 91-110). New York, NY: Longman.
- Rowe, W E.. (2002). *A meta-analysis of six Washington State restorative justice projects*. For July 2000-June 2001. Olympia, WA: Governor's Juvenile Justice Advisory Committee (GJJAC), Office of Juvenile Justice.
- Roy, S. (1994). Victim offender reconciliation program for juveniles in Elkhart County, Indiana: An exploratory study. *Justice Professional*, 8(2):23-35.
- Sherman, L. W. (1993). Defiance, deterrence and irrelevance: A theory of the criminal sanction. *Journal of Research in Crime and Delinquency*, 30, 445-473.
- Sherman, L., Braithwaite, J., Christie-Johnston, J., Inkpen, N., Smith, S., & Strang, H. (1997). Reintegrative shaming of violence, drink driving & property crime: A randomized controlled trial. Progress report. Canberra, AU: Australian National University, Institute of Advanced Studies, Research School of Social Sciences.
- Sherman, L., Gottfredson, D., McKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising*. Washington, DC: Office of Justice Programs.
- Sherman, L. W., Strang, H., & Woods, D. (2000). Recidivism patterns in the Canberra Reintegrative Shaming Experiments (RISE).. Centre for Restorative Justice. Research School of Social Sciences. Australian National University.
- Strang, H. (2001). Justice for victims of young offenders: The centrality of emotional harm and restoration. In G. Maxwell & A. Morris (Eds.), *Restorative justice for juveniles* (pp. 183-193). Oxford: Hart Publishing.
- Strang H. (2002). *Repair or revenge: Victims and restorative justice*. Oxford and New York. Oxford University Press.
- Sundell, K., & Vinnerljung, B. (2004). Outcomes of family group conferencing in Sweden: A 3-year follow-up. *Child Abuse & Neglect*, 28: 267-287.

- Trimboli, L. (2000). An evaluation of the NSW Youth Justice Conferencing scheme. Sydney, Australia: New South Wales Bureau of Crime Statistics and Research.
- Tyler, T. R. (1990). *Why people obey the law*. New Haven, CT: Yale University Press.
- Umbreit, M. (1994). *Victim meets offender: The impact of restorative justice and mediation*. Monsey, New York: Willow Tree Press.
- Umbreit, M., & Roberts, A (1994). *Mediation of criminal conflict in England: An assessment of services in Coventry and Leeds*. St. Paul, MN: Center for Restorative Justice & Mediation. University of Minnesota.
- Umbreit, M., Coates, R., & Vos, B. (2000). *Victim impact of restorative justice conferencing with juvenile offenders: What we have learned from two decades of victim-offender dialogue through mediation and conferencing*. Boca Raton, Florida: Florida Atlantic University, Community Justice Institute, Balanced and Restorative Justice Project.
- Weibe, R. (2003). Reconciling psychopathy and low self-control. *Justice Quarterly*, 20, 297-336.
- Wilson, R. & Prinzo, M. (2001). Circles of support: A restorative justice initiative. In M. H. Miner & E. Coleman (Eds.), *Sex offender treatment: Accomplishments, challenges, and future directions* (pp. 59-77). Binghamton, NY: Haworth Press.