From Evaluation to Interpretation

Assessing Theoretical Elements in Residential Programmes

CHRIS BAERVELDT
Utrecht University, the Netherlands

BOB HORJUS
Researcher/consultant, the Netherlands

MICHA DE WINTER
Utrecht University, the Netherlands

Impact studies of prevention programmes, in particular meta-analyses, usually interpret outcome and impact statistics as tests of an underlying theory of prevention. However, these programmes usually combine various interventions linked to different theoretical perspectives. Consequently, the effects of the programme can easily be misinterpreted. This article introduces an interpretation method that acknowledges the eclecticism of prevention practice and is also a feasible instrument to enhance the quality of meta-studies. First, the interventions of the programme are identified and analysed separately. Second, an assessment is made of the arguments that link interventions and the espoused theories. Third, each intervention is represented by a set of scores indicating the types of links and core theoretical assumptions. These scores are aggregated to programme scores and included as independent variables in meta-analyses. The method is illustrated by an evaluation of practices in two Dutch crime prevention institutions. The evaluation demonstrated the theoretical value of (eclectic) practices, even when they differ substantially from the official programme theories. The approach also highlighted an 'interpretation error' in assessing the impact of one of the programmes and suggests a correction.

KEYWORDS: adolescents; evaluation; impact; interventions; prevention; theory

Introduction

Modern western society invests large amounts of money in trying to influence the behaviour of juvenile offenders. Currently, the USA spends more than $20 billion

per year on correctional youth institutions alone, and in 1996 it spent $3.2 billion on youth crime prevention programmes (Sherman et al., 1997). In a small country like the Netherlands, the government spends the considerable sum of 285.2 million euros on young offender institutions (Donner, 2004). Accordingly, many impact evaluations have been carried out since the 1960s (Rossi et al., 1999); for instance, in a meta-study of juvenile crime prevention programmes in the USA, Lipsey (1995) reported about 400 studies. The rationale behind these impact studies is that intervention designers learn from the past and will drop or modify programmes that are proven to be ineffective. According to Lipsey et al. (2000), designers might try other programmes, or at least make changes in some of the standard programmes – deterrence programmes, survival and challenge programmes and parole – when these seem to have no positive effects on recidivism. Sometimes they use a list of interventions that are believed to be proven successful, and they only design new interventions based on this list (see, for a Dutch example, Ince et al., 2004). It is argued here that undertaking impact evaluations of earlier programmes to predict the success of intervention programmes can be unreliable. Our argument is not that evaluations of effects often lack technical quality (Cook and Campbell, 1979), because there are many studies that do meet such quality criteria, but that effects can easily be misinterpreted.

In their classic study, Cook and Campbell (1979) review quasi-experimental research designs, their validity problems and solutions to these problems. Typical impact studies investigate the effect of an independent variable X on a dependent variable Y. The dependent variable Y represents the outcome, for instance the delinquency level of adolescents. Variable X refers to the content of a programme – for instance, social skills training, whereby the training is assumed to improve social skills and in turn, that better social skills will reduce delinquency. If social skills training reduces delinquency, the programme is said to be successful. Consequently, effects of X on Y are considered tests of a general theoretical hypothesis, in this case of the effect of social ability (or learning) on delinquency level. This approach is usually followed in meta-analyses (Boendermaker, 1999; Lipsey, 1995; Thalheimer and Cook, 2002). Programmes are evaluated and compared by effect sizes of X on Y, while statistically controlling for their reliability. Large and statistically significant effect sizes indicate successful programmes. As a consequence, similar programmes, i.e. programmes using the 'same' X, are also predicted to be effective.

However, the interpretation of the effects of X on Y is hazardous, because many programmes include more than one type of intervention. Prevention programmes usually tend towards eclecticism: they include all kinds of interventions including interventions that do not seem to fit the main programme design. A social learning programme, for example, may include social training sessions as expected, but also other interventions such as family therapy. When this programme has been proven successful in an impact study, this might as easily be attributed to social learning as to family therapy, and future programme designers do not have an empirical basis to choose between the two. According to Sherman et al. (1997) 'the problem is that even with successful results, a combination . . . makes it impossible as a matter of scientific method to isolate the active ingredients causing the success. It may be all of them in combination. Or it may be only one or two.'

Researchers have reacted differently to this problem, ranging from ignoring it to giving up any attempt at theory-based interpretation. This article proposes a middle road between these two options, describing each programme with a small (e.g. between 3 and 7) number of independent variables: $X_1, \ldots, X_n$. This would increase the number of independent variables in meta-studies, but keep the number small enough to make analysis feasible. It would also be more in keeping with (the reality of) eclectic practices. However it is not necessary to give up hope that knowledge from impact studies can be generalized and used for the design of future programmes. The set of variables $\{ X_1, \ldots, X_n \}$ has to be meaningful, i.e. a description of programmes by such a set should include the aspects thought by evaluators and practitioners to be essential. This article shows how such sets of variables can be designed and used to interpret programmes in young offender institutions, illustrated by presenting an evaluation of programmes in two Dutch young offender institutions.

Interpretation Problems and Solutions

The literature illustrates four different approaches to tackling the interpretation of impact evaluation findings and the problems this presents. The first is to ignore the problem, which is often the response in meta-evaluations. These studies usually rely on official descriptions of the programmes to be evaluated, which often do not mention interventions that do not fit neatly within the supposed programme design. Each programme is placed within only one category of the meta-evaluation, even when the programme apparently fits into more categories. For instance, Lipsey et al. (2000) categorized an intervention described as 'a 12-week cognitive mediation training programme involving small discussion groups' as an example of 'behavioural programmes' and ignored the fact that the programme developers designed group dynamics in their programme. In general, as Lipsey and Wilson (1998) showed in a meta-analysis of interventions for serious juvenile offenders, the content of the programmes classified as belonging to the same category may vary widely. As has been suggested, such broad classifications may lead to poor interpretations of effects. Lipsey and Wilson report that the effects of programmes classified as belonging to the same category differed substantially. They also show that variables that probably indicate different 'treatment' content in programmes, like the integrity of treatment implementation, the programme's age and the qualification of personnel, had strong additional effects on programme outcomes.

The second approach to this problem is to shape the actual practice in programmes to strictly meet theoretical standards. This often includes a list of interventions that are minutely described, as well as strongly formalized and controlled intervention processes. Medical research usually stresses the methodological necessity of treatment integrity, also called 'treatment fidelity' (Nigg et al., 2002), which Moncher and Prinz (1991) define as 'the confirmation that the manipulation of the independent variable occurred as planned'. Treatment fidelity might be a feasible strategy in controlled medical experiments, but in most social intervention programmes it is an illusion. For instance, Marks (2002)

states that therapists 'use a bit of this, a lot of that and none at all of the third component of the list. Yet all therapists call their method cognitive-behavioural therapy.' The main reason for this practical eclecticism is that evaluators and interveners have different objectives: the first want to test one general theory, while the latter want to 'cure people' and to this end will use interventions that may or may not be compatible with theory. While theoretical purity in social intervention programmes might be desirable from a researcher's perspective, it is illusory in practice (see also Michie and Abraham, 2004).

The third approach might be characterized as 'practical empiricism'. Many researchers reacted to the interpretation problem by viewing impact studies as unique enterprises around unique programmes (Lipsey, 2001). Instead of describing programmes by one variable X, as is done in most meta-studies, these researchers seem to describe them by large and ever growing numbers of variables – the result is a practically infinite number of different types of programmes. This approach can be criticized because it does not seem to add to the knowledge that future intervention designers might need. Some researchers (e.g. Damen and Delicat, 2004) have tried to achieve generalized knowledge by a detailed systematic analysis of case studies, including minute descriptions of goals, methods and actual practices of interventions and their relationship with changing behaviours and attitudes. However, it is still hard to generalize from this kind of study, because their sense of detail is so strong that comparisons across projects or activities are difficult. As a consequence, the practical value of their studies for intervention design is limited.

The fourth approach includes studying components of programmes. Authors like Marks (2002) propose to 'dismantle' programmes into treatment components that can be analysed separately. A feasible way to do this would be to make a list of all interventions within a programme and analyse them. This approach has the merit that it captures a substantial part of the complexity: that programmes may include all kinds of interventions, and the list of interventions might vary widely between versions of 'the same' programme in one institution and another. Therefore, this article follows Marks's suggestion, and will analyse programmes by analysing their interventions separately. Yet two problems of this approach remain. The first is how to analyse interventions. The second problem is how to aggregate the knowledge about separate interventions into some final programme profile. This article, meant as an introduction to outline a new method, focuses on the first problem, while only briefly going into the second which is more technical. Regarding analysis of the interventions, we note that there is an abundance of different interventions in the field, and most interventions differ at least somewhat from others. Consequently, the straightforward approach to identify the major intervention components or themes, e.g. whether a social skills training component or group meetings are included, would result in a long and seemingly endless list. This would lead to a large number of dummy variables in (meta-)analyses, too large to make such an analysis possible or fruitful. Therefore, this number has to be reduced. This was undertaken by categorizing the interventions. While seemingly easy at first sight, it is quite complicated because theoretical assumptions may influence the categorization system: that which is assumed to

be an important feature of an intervention can vary from one theoretical perspective to another. Usually, there is no compelling reason to keep to only one or a few theories when classifying interventions. In effect, when designing the categorization system, many evaluators would want to have an overview of all possibilities from all possible theoretical viewpoints first, before choosing a certain approach. Therefore, the main task here is to design a categorization system that is open to insights from all (or nearly all) theories in the field, whilst still keeping the number of variables low.

Interventions and Theory

According to Cook and Campbell (1979) programmes are directed at the change in an outcome variable Y. Suppose we compare only programmes with the same Y, for instance delinquency. We propose that (meta-)analysts should use meaningful and relatively small sets $\{X_1, \ldots, X_n\}$ of independent variables to describe programmes instead of the classical X that only indicates whether a programme is, for instance, a social learning programme or another type of programme. This article focuses on meaningful variables reflecting the content of interventions. However, we acknowledge that programme effects are not likely to be the outcome of the content of interventions alone, but also, and sometimes even more, of such variables as the programme's budget, staff and management quality, embeddedness in the environment and the type of client group (e.g. Landenberger and Lipsey, 2005; Lipsey and Wilson, 1998). Many effects of these 'non-content' variables are reflected in the effects of content variables. For instance, when the budget is small or staff and management quality is low, this will be reflected in a poor repertoire of interventions, which will often weaken the content of the programme. Note that including the $\{X_1, \ldots, X_n\}$ instead of the single X can enlarge the variance of Y explained by independent variables that derive from the content of intervention programmes. However, there is probably still a separate effect of non-content variables left. For instance, one can expect the programme's embeddedness and type of client group to have effects on Y apart from the effects on the content of interventions. Thus, we propose to include at least some of these non-content variables in the analysis.

Linking Interventions to Theory

We propose to analyse the interventions within the programmes separately and then to aggregate the results. We also propose to classify interventions by their links to theory. But what is a link and how to assess it? As has been discussed, the relationship between interventions and theory is much looser than is suggested by most programme evaluations and meta-analyses. Programmes usually deviate from the programme description, as do the interventions within the programme. In fact, most programmes and interventions are not explicitly designed from a theoretical perspective at all, but seem to grow from what is available and acceptable in the field. However, while most professionals in the field do not refer explicitly to theory, their interventions still *can* be traced back to some theory.

Consider, for instance, family counselling, seen as a means to prevent delinquent behaviour. Interveners do not choose family counselling arbitrarily; even if they cannot formulate the reasons, there are still arguments for its use. These arguments have a theoretical value: even though professional interveners sometimes are not able to mention the principal authors of the theories, they have at least some general idea about the most important assumptions. Family counselling could be argued to be an effective means for the following reasons: (i) counselling improves family bonds; and (ii) the improvement of family bonds reduces delinquent behaviour. The latter argument can be traced back to, among others, social control theory (Hirschi, 1969). Consequently there is a link between family counselling and social control theory.

Note that it is not essential for practitioners involved in an intervention to actually employ theoretical arguments to justify their interventions. They may use family counselling without 'acknowledging' any such arguments. It is, however, critical that they *could* use these arguments. Note also that the link between family counselling and social control theory does not mean that the argument should refer to all aspects and details of social control theory. Social control theory includes other assumptions, such as:

- people have a tendency to delinquent behaviour because it would be rewarding when no controls were in place;
- bonding relates to the domain of the family but also to other domains such as school; and
- the effect of bonding on delinquency is universal and constant across all people.

These other assumptions are not required to justify the use of family counselling. In general, interventions are not justified by referring to entire theories, but only by reference to one or some of the core assumptions in those theories.

Logical Links

By definition what we have labelled as a 'logical link' – the linkage between an intervention and a theory – exists when the use of the intervention can be justified by reference to at least one core assumption of the theory. More precisely, a link exists when two conditions are met: (i) the intervention increases or decreases the value or rating of the target population in relation to a certain (set of) variable(s); (ii) it can be concluded from the core assumption that such a change also influences the target behaviour as intended. For instance, a logical link exists between family counselling and the core assumption that strong bonds prevent delinquency, when it is true that (i) family counselling increases the strength of the bonds of the target group, and (ii) it can be predicted from the core assumption that increasing the bonds of the target group also decreases delinquency.

Note that according to this strict definition, logical links do or do not exist. However, in practice the existence of a link is often better expressed as a probability. While professionals usually agree about the second part of the link, i.e. what predictions about variable change can be inferred from core assumptions, they

often disagree about whether certain interventions actually change those variables. They may rate these effects in terms of probability. Therefore, we usually assess the probability of logical links instead of their existence. However, for the sake of argument, in this section logical links are referred to as existing or not.

Note that an intervention can have logical links to different core assumptions. Family counselling could also be linked to the core assumption that self-efficacy decreases delinquent behaviour. Assuming that family counselling enhances self-efficacy, it would also decrease the probability of delinquent behaviour. Interventions have logical links to certain theories, and not to others. For instance, the use of family counselling usually cannot be justified convincingly with reference to core assumptions of an opportunity theory of delinquent behaviour or a relative deprivation theory. But job search assistance could because it would decrease relative deprivation. Consequently, different interventions often have different sets of logical links. They only have the same sets when they can be justified by the same arguments. Thus, interventions with logical links to the same set of theories have common features that set them apart from other interventions. We are therefore able to label interventions theoretically with much more nuance than in classical meta-evaluations, where interventions, and even entire programmes, are linked to only one theory.

Theoretical Elements

While we are now able to categorize interventions by their logical links, we cannot use these links in meta-analyses, because the total number of core assumptions in relevant theories would be too large; and large numbers of independent variables would prevent statistical significance of parameters in most meta-studies. Therefore, we propose to classify core assumptions into broader categories in such a way that: (1) the number of categories is relatively small (e.g. between three and seven); (2) (nearly) all theoretical core assumptions can be categorized; (3) (nearly) all interventions can be categorized. We will call these categories 'theoretical elements', because they refer to classes of theoretical core assumptions that might be common to different theories.

Let us assume that we have a set $\{X_1, \ldots, X_n\}$ of theoretical elements that meet these criteria. Then, we can classify interventions by the existence of theories linked to this set. By definition, an intervention has a logical link to a theoretical element X_i when it has a logical link to at least one of the theoretical core assumptions that are included in X_i. Consider, for instance, an X_i which is defined as the class of core assumptions which predict that delinquency will decrease because the social costs are higher. One of the core assumptions that falls into this class is the one that family bonds prevent delinquency. Now, family counselling has a logical link to X_i because it has a logical link to the core assumption that family bonds prevent delinquency, the counselling makes the bonds stronger and delinquency decreases when bonds are stronger. Thus, all interventions can be classified according to the existence (or probability) of logical links to all theoretical elements. Consequently each intervention has values on a set of variables $\{X_1, \ldots X_n\}$ corresponding with the elements.

It is difficult to design a small set that is not blind to some important theoretical arguments of certain theories and we have often switched between possibilities and variants before coming up with useful sets. In our experience it works well to start from a practice-inspired theoretical approach already so eclectic that it is almost impossible to falsify its hypotheses. We used such approaches, 'stripped' them roughly to elements, and turned them into useful sets in several fields. However, rather than trying to explain the process of designing in rather awkward abstract terms, we would like to present two examples. First, we look at the set of theoretical elements that was defined by Baerveldt (1993) in order to evaluate the effects of 200 local crime prevention programmes in the Netherlands. These programmes varied widely with respect to the target behaviour, the target (client) group and the intervention methods that were used. The programmes included, for instance, investing in better hardware against burglary, social work projects for adolescents, neighbourhood watch and school attendance control systems (Polder and Van Vlaardingen, 1992). Many of these interventions could be (and sometimes even were) justified with reference to core assumptions from different criminological theories, like stress–strain theory, social control theory, differential association theory and opportunity theories. According to Baerveldt (1993), these core assumptions could be classified into four categories. The first element X_1 included core assumptions indicating that behaviour is caused or enhanced by social rewards, like esteem, social capital or emotional support. In contrast, the second element X_2 referred to social costs, for instance the effect of social sanctions or isolation on behaviour. The third element X_3 included all core assumptions referring to non-social rewards, such as the job satisfaction of burglars. The fourth element X_4 likewise referred to non-social costs, such as time, labour and legal sanctions. The set $\{X_1, X_2, X_3, X_4\}$ meets the three criteria for a useful (meaningful) set. First, the number of elements, four, is small enough to keep the number of independent variables in a meta-study manageable. Second, the core assumptions of (almost) all relevant criminological theories can be classified. For instance, social control theory predicts less delinquency when the social costs are higher, but ignores other elements (note that Hirschi (1969) explicitly indicates social control theory to be a theory of social costs). Differential association theory (Sutherland, 1947) predicts delinquency levels only from the social rewards and social costs that persons expect from their friends, and ignores non-social rewards and costs. Classical strain theory predicts delinquency levels only from non-social rewards, in particular monetary value, and opportunity theories usually ignore social rewards and costs, but predict delinquency from non-social rewards, non-social costs or both. Third, nearly all interventions can be categorized with respect to their logical links to the four elements. This was done by Polder and Van Vlaardingen (1992), who asked a team of specialists to rate all interventions as described in programme evaluations. Family counselling, for example, was rated as linked to X_1 (social rewards), X_2 (social costs), but not to X_3 and X_4 (non-social rewards and non-social costs). Job search was rated to have a logical link to X_3 (non-social rewards), and in some cases, also with X_1 and X_2 (social rewards and social costs).

Using Theoretical Elements in Meta-Analyses

A useful set $\{X_1, \ldots, X_n\}$ of theoretical elements enables researchers to rate all interventions. The rating system might be more flexible than just indicating whether a logical link with a certain X_i exists ($X_i = 1$) or not ($X_i = 0$), and use a probability score: $0 \leqslant X_i \leqslant 1$. Whatever rating system is used, the rates of the interventions should be aggregated to the level of programmes before the analysis starts. The values of a prevention programme on $\{X_1, \ldots, X_n\}$ are aggregated from the values of the corresponding X_i of all the interventions within the programme. Values can be aggregated by computing means or maxima or other functions. The aggregated values can now be used as independent variables in meta-analyses. The model for the analysis might have the next form:

$$Y = \beta_0 + \beta_1 \times_1 + \ldots + \beta_n \times_n + \Sigma_{\text{other}} + \varepsilon$$

where Y is the outcome variable, X_i represents the scores on the theoretical elements, and Σ_{other} a set of important non-content variables regarding, for instance, target group, management and the history of the programme. Unfortunately, the quality of most programme evaluations of the Dutch example was far below standard, and thus, the number of reliable effect studies was too small for a meta-analysis (Polder and Van Vlaardingen, 1992).

Theoretical Elements and Residential Crime Prevention Programmes

A method has been outlined to improve the interpretation of effect evaluations by substituting the classical single X indicating that a programme is of a certain type, for a small set of X_i that is more informative about the content of programme interventions. The concept of logical links between interventions and theoretical core assumptions was introduced, and it was argued that classifications of the core arguments, i.e. theoretical elements, can define feasible sets of X_i. An example was employed to demonstrate that it is possible to define a useful set of theoretical elements for the analysis of a wide range of interventions against delinquency. In the remainder of this article we will introduce a set that was especially designed for the analysis of residential crime prevention programmes, and show how it was used in the analysis of the programmes of two Dutch youth offender institutions. Thus, we show how sets can be designed according to common features of investigated projects. Moreover, we also show how logical links between the interventions and theoretical elements might be assessed, and how to aggregate from the logical links of the interventions to the X_i values of the programme.

    The Dutch juvenile crime prevention system includes all kinds of institutions, some private, some run by the state, which offer a large variety of prevention programmes for adolescents. Some provide short-term care and others offer an extensive programme of rehabilitation. What almost all of these prevention programmes have in common is that their objective is to prevent adolescent recidivism after

leaving the institution, and that they try to accomplish a 'proximal effect' (Rossi et al., 1999) by changing the adolescent's behaviour or values in institutional settings. They receive or are exposed to social skills training in the institution, because it is assumed that better social skills will lead to better conflict management and relationships after their release, which will prevent delinquent or other anti-social behaviour. Therefore, one of the most important issues of these programmes is whether the adolescents will keep and use their new capabilities when they re-enter the outside world (e.g. Boendermaker, 1999; Rossi et al., 1999). The set of theoretical elements introduced earlier was designed for community crime prevention programmes, and ignores this issue. For the analysis of residential programmes we defined another set of theoretical elements, which we called the threshold set.

Theoretical Elements: The Threshold Set

The threshold set consists of four elements, namely the threshold element, the attitude element, the social norms element and the self-efficacy element. The first element X_1, called the threshold element, refers to differences between the situation inside the institution and after release. The common core argument here is that when such differences are larger, the association between proximate behaviour (and intentions) at the institution on the one hand, and the target behaviour after leaving the institution on the other hand, is smaller. Now, by definition, an intervention has a logical link to the thresholds element when the intervention, for instance job search or social aftercare, can be argued to bridge the gap between proximate and target behaviour. The second element is called the attitude element. The common core argument here is that positive proximate behaviour and intentions follow from positive attitudes regarding the target behaviour. An intervention has a logical link with the attitude element when the intervention makes attitudes more positive, and consequently also the proximate behaviour. Similarly, a link between an intervention and the social norms element exists when the intervention changes the clients' perceptions of the norms of their social environment (including, e.g. family, peers and neighbourhood), and consequently the proximate behaviour improves. Finally, a logical link between an intervention and the self-efficacy element exists when the intervention changes the inmate's ability to change his/her behaviour as intended.

As we have already argued, a set of theoretical elements is useful when (1) the number of categories is small enough (e.g. between three and seven) to be used in meta-analyses; (2) (nearly) all theoretical core assumptions can be categorized; (3) (nearly) all interventions can be categorized. The first criterion is easily met, because the set includes four elements. With respect to the second criterion, we would like to stress that the set is purposefully designed to bear a close resemblance to the central concepts used in the common literature with respect to the relationship between intentions and actual behaviour. In particular, we refer to Ajzen's (1991) theory of planned behaviour, and its many offspring and variants. According to Sutton (2003), the theory of planned behaviour covers all possible direct effects on behaviour; and other effects (e.g. of sociodemographic, cultural and personality characteristics) can only be explained by their influence on the

causal (independent) terms of the theory of planned behaviour. In fact, most theories in the field include core arguments that can be classified by one or two elements of the set. For instance, social control theory agrees with the social norms element, and social learning theory agrees with both the social norms and the self-efficacy element. With respect to the third criterion, we would like to remark that the theory of planned behaviour and its offspring have become highly popular in the field, and its concepts are often used as checklists in practice. Therefore, we may expect that the main part of interventions can be justified with reference to at least one of the elements. Most interventions, like multi-component group treatment (Leeman et al., 1993; Nas et al., 2005) can easily be motivated with reference to one of the elements (in this case: the attitude element).

Using the Threshold Set: Two Residential Crime Prevention Programmes

We used the threshold set to study the programmes of two Dutch young offender institutions. The first institution, the Hoenderloogroep (1999a, 1999b) started a Glen Mills high school in a former military training centre. The core of the Glen Mills approach includes the manipulation of group dynamics by an inmate status system. The achievement of within-group status is thought to be extremely important for members of delinquent groups. Therefore, this system has been adopted in the institution, naturally with status indicators that are meant to encourage non-delinquent behaviour. Boys entering the school have the lowest status, including a lack of privileges and a strict regime. They can gain status and rights by desirable behaviour, such as good school results or pro-social behaviour. The status level is not authorized by staff but by inmates with a high status. The inmates can lose status as easily as they gain it. Higher status inmates have some influence on staff decisions. The staff is expected to facilitate only structural conditions like a strict regime of activities, schooling, employment training and active mediation in finding a job. The status system is self-regulatory, as, according to the theory, the inmates keep it running and thus are motivated to abide by the regime and invest in school. Note that the arguments underlying the programme typically refer to the social norms element: the inmates are expected to change their behaviour due to a system of social pressures and rewards, in which the interventions are used to increase that pressure. There is some reference to the self-efficacy element, because some of the interventions are also designed to increase self-efficacy. However, the programme is not designed to change attitudes as a means to change the inmates' behaviour: attitudes might change, and are expected to change in the long run, but only because the inmates' behaviour changes first. Likewise, the programme is not designed to decrease the threshold between the behaviour in the institution and the behaviour after release. In practice, however, Glenn Mills schools might deviate from the designed model and use additional interventions that can be traced back to other elements than the expected ones. Therefore, it is useful to analyse the interventions used in practice.

The second institution, a youth prison called the Kolkemate, has been using a competence achievement model as its official programme since the second half

of the 1990s (Spanjaard, 1998). According to the institution's working theory, adolescents have to complete developmental tasks like the achievement of a personal and relational identity, of professional knowledge and skills and the acceptance of societal rules (Slot, 1999). The fact that the inmates have not been able to complete these tasks on their own is considered the cause of their problems. Accordingly, the institution's goal is to teach them how to do so. Because there are many developmental tasks in adolescence, training activities also encompass a wide range of topics, including, for instance, daily routine activities like time-keeping or personal hygiene, as well as social skills training. While the choice of topics reflects the personal goals of the inmates, there is a firm control of the inmates' participation. Note that the arguments underlying the programme typically refer to the self-efficacy element: the inmates are expected to change their behaviour due to a system of improvement of skills, and the interventions are expected to improve these skills. The other elements do not play a role here. The competence achievement model does not include learning from the social norms of other group members. Also, attitude change is not considered a means to change the behaviour of inmates. Finally, the model's design does not include interventions aimed at decreasing the threshold between behaviour inside and outside the institution. Again, the institution might deviate from the design and include interventions that have links to other elements than self-efficacy.

We evaluated the logical links between the interventions and the theoretical elements in both institutions. We made a list of interventions, each with short descriptions from reports of or interviews with staff members, including 13 interventions in the Hoenderloogroep and 20 in the Kolkemate (see Tables 2 and 3). Four field specialists, with wide experience as prevention worker or adviser in adolescent welfare or juvenile delinquency, were asked to rate the logical links. They only received the list of interventions and short descriptions but were not informed from which institutions or programmes the interventions came. We asked the specialists to rate whether and how much each intervention affected the threshold between intention and behaviour, and the attitude, social norms or self-efficacy of the inmates. The link of each intervention to each element was rated by a score between −4 and +4. A score of +4 indicated that the intervention could be expected to have a strong positive effect on the element. A zero score indicated no effect at all, and negative scores (to a minimum of −4) indicated counter (negative) effects. At first sight, the scores were not reliable, because the specialists' mean scores differed systematically. However, these differences decreased substantially when the scores were aggregated to the level of programmes. Table 1 shows that, even while rater 2 rates much higher than the others, all raters evaluate the links in the Hoenderloogroep in exactly the same order: the links to the social norms elements are strongest, followed by the effects on self-efficacy, attitudes and thresholds respectively. Likewise, the effects in the Kolkemate on thresholds are rated as the weakest by all raters. The effects on social norms are rated largest by raters 2 and 4. Raters 1 and 3 rated them second, but their second scores were practically the same as their largest. The rates on attitudes and self-efficacy are in between, with some order differences between the raters.

Table 1. Mean Strength of Logical Links of Interventions in the Hoenderloogroep and the Kolkemate with Theoretical Elements of the Threshold Set $X_1, ..., X_4$

Theoretical element			Rater/specialist				Mean score
			1	2	3	4	
Hoenderloogroep							
	X_1	Threshold	.08	2.50	.21	−.71	.52
	X_2	Attitude	2.07	2.64	1.82	2.64	2.29
	X_3	Social norms	2.43	3.36	2.43	3.43	2.91
	X_4	Efficacy	2.21	2.93	2.36	2.93	2.61
Kolkemate							
	X_1	Threshold	−.11	1.47	.68	−.06	.50
	X_2	Attitude	2.08	2.05	1.18	1.78	1.77
	X_3	Social norms	2.05	2.37	1.47	2.95	2.21
	X_4	Efficacy	1.63	2.26	1.58	2.42	1.97

4 = maximum score, indicating a strong link; a zero indicates no link; negative scores indicate negative effects. Rating scores by four field specialists.

Table 1 shows that the logical links of the interventions in the Hoenderloogroep largely reflect what could be expected from a Glenn Mills programme: the links to the social norms elements were the strongest. The links to the self-efficacy element were second and the links to the threshold element were weak, both as expected. However, some of the interventions were also rated as affecting the attitudes of the inmates. The logical links of the interventions in the Kolkemate deviated strongly from what was expected from the official programme description. Link to self-efficacy was expected to be strongest, but was in effect rather weak and weaker than the link to social norms, which was expected to be weak. The programme did not seem to contain interventions that could be justified by any core assumption. All specialists consistently rated the logical links of the institution's interventions lower with respect to self-efficacy, attitude and social norms than those of the Hoenderloogroep, even though the list of interventions was different and the raters could not relate any of the interventions to an institute.

Conclusions and Discussion

In this article we propose a method of interpreting the content of crime prevention programmes that acknowledges the eclecticism of practical prevention methods and is still useful for meta-analyses. For practical reasons, we have not yet been able to show our method in a meta-analysis, nonetheless these examples showed that it is possible to design *sets* of elements for community and residential crime prevention programmes. We also showed that the latter set can be used for the analysis of the programme of two young offender institutions.

Table 2.   Interventions in Young Offender Institution 1: The Hoenderloogroep

Living in an extremely structured surrounding	Every hour of the week is planned and the inmates have to keep to their schedules.
Individual mentoring	The inmates have regular talks with their appointed mentors.
Sports	The inmates participate in sports activities on a daily basis.
Attending school	Each inmate follows a schooling programme on the premises in which he can obtain a regular diploma.
Staying on a voluntary basis	The youngsters can choose between internment in the Glen Mills School or in a youth detention centre. If they wish to leave, they may, but this results in internment in a detention centre.
Being tried by fellow inmates	Each inmate must face a tribunal of fellow inmates when he has broken a rule.
Being under group influence	The inmates live in groups and are constantly under the control of fellow inmates.
Promoting hierarchy among inmates	There is an strict hierarchy among the inmates representing different stages and privileges. Each ascent in the hierarchy is based on an improvement in behaviour.
Privileges for top group of inmates	A top group of inmates is permitted to take decisions, give orders and represent the institution.
Attending group meetings on a daily basis	In daily group meetings the inmates receive instructions and have the opportunity to voice their opinions.
Learning about norms of conduct	Inmates are taught how to behave according to the norms of the institution and society.
Learning social and practical skills	Each inmate follows a group training programme of social and practical domestic skills.
Working and earning wages	Members of the top group are permitted to work outside the institution in a regular job.

The practices in these institutions deviated from what could be expected from the official programme design. One institution, the Kolkemate, seemed to have serious problems with establishing its programme as intended. The other one, the Hoenderloogroep, had incorporated interventions directed at attitude change. These interventions were not mentioned in the official programme. While there is no practical argument against such 'additional' values in a programme, and practical eclecticism might even be more an asset than a drawback, it is problematic for researchers. An affirmed success of the programme, as suggested by Van den Bogaart et al. (2003), could easily lead to an incorrect interpretation, namely that programmes which are only directed at social norms and self-efficacy are effective. By using the threshold set, we were able to point out an interpretation error, and suggest a correction.

We asked four specialists to rate the logical links between the interventions and the theoretical elements. We feel that this method is better than previous practices in meta-studies, usually going no further than copying the official programme description or the 'educated guess' of (only) researchers about programme content. However, even while there was consistency between the raters

Table 3.   Interventions in Young Offender Institution 2: The Kolkemate

[Being subject of planned skill training] Goal orientation	The institution has a goal oriented programme in which inmates are trained in setting goals for themselves and how to achieve these.
Learning social skills	Inmates receive group training in how to behave and solve interpersonal problems.
Obligatory participation in the programme	The daily programme is not voluntary. All inmates must participate.
Attending school	Each inmate follows a schooling programme on the premises in which he can obtain a regular diploma.
Living in a residential institution	The inmates live together in a residential institution.
Individual mentoring	The inmates have regular talks with their appointed mentors.
Token economy system	A system of reward tokens serves to grant privileges and monitor the inmate's behaviour.
Appropriate working relation with the personal mentor	The relation with the mentor is considered an aspect of the intervention. The mentor is a role model and helps with the training process.
Sports	The inmates participate in sports activities on a daily basis.
Reinforcing the personal network	Care is taken to make contact with the personal network of the inmate to ensure support for the inmate in the rehabilitation process.
Preparation on the situation after release	The inmate receives training aimed at improving his resettlement in society.
Attending group meetings	In regularly held group meetings the inmates receive instructions and have the opportunity to voice their opinions.
Phased extension of leave	Depending on the length of his sentence and his progress in the programme the inmate will be granted progressively longer periods of absence.
Aftercare	In the rehabilitation process the youngsters receive help and counselling.
Temporary expulsion in cases of misconduct	In cases of extreme misconduct the inmates are temporarily separated from the group.
Behavioural instructions	Inmates are individually taught how to behave in a more socially acceptable manner.
Living in confinement	Inmates experience total confinement.
Analysis of the inmate's crime	In talks with the mentor the inmate analyses his crime. This analysis is aimed at making him aware of his crime and helping prevent him from reoffending.
Strict schedule	Every hour of the week is planned and the inmates have to keep to their schedules.
Training in Japanese sports	The inmates get training in Japanese martial arts.

regarding the rank of the scores of the programmes, we think that the inter-rater consistency per intervention should be improved. This is certainly possible: in an ongoing study we were able to improve the inter-rater consistency substantially by a group training using examples of descriptions. This also led to better intervention descriptions. If time, money and legal possibilities permit, researchers can improve their descriptions by better observations and use of the institution's protocols and other materials, as well as by interviewing inmates. The rating could improve by using a standard format for descriptions. Such a format should systematically not only include goals and means of the intervention, but also a list of involved interveners, their time budget, and the amount of time inmates are confronted with the intervention.

We aggregated the programme's values over the X_i by simply computing the mean scores over all interventions. Other aggregation methods could also be used. We computed the maximum scores and found similar results. It could be argued, however, that more sophisticated methods are needed when there are certain logical connections between interventions; for instance, when one intervention only follows after another. In that case more in-depth knowledge of the programme is necessary.

We stressed that we wanted to improve meta-analyses. We acknowledge that the best way to do so is to present a better one. In an earlier stage, we analysed the data of a Dutch study on community crime prevention programmes (Polder and Van Vlaardingen, 1992). Unfortunately, we were not able to assess any significant effects because the number of reliable effect studies was too small for a meta-analysis. Therefore, we have not yet been able to show the value of our approach in a meta-analysis. However, it is easy to imagine how such an analysis should be carried out. We suggest that this can be achieved by demonstrating relationships between sets of intervention and outcome variables linked by threshold variables that elaborate on the logical links embodied in programme practice. Thus a future model might be stated as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \Sigma_{other} + \varepsilon$$

where Y represents the outcome variable, X_1 reflects the strength of the logical links of intervention programmes with the threshold element, X_2 with attitude, X_3 with social norms and X_4 with self-efficacy. Thinking in terms of such theoretical elements the theoretical interpretation of effects offers a better chance of accurate interpretation of outcomes than in classic impact studies, and hence offers programme designers a better possibility for drawing accurate lessons from past interventions.

References

Ajzen, I. (1991) 'The Theory of Planned Behavior', *Organizational Behavior and Human Decision Processes* 50: 179–211.

Baerveldt, C. (1993) 'Het Gebruik Van Rationele Keuzereconstructies Bij De Evaluatie Van Interventies', *Tijdschrift voor criminologie* 35(2): 158–76.

Boendermaker, L. (1999) *Justitiële Behandelinrichtingen Voor Jongeren. Populatie En Werkwijze*. Leuven and Apeldoorn: Garant.

Cook, T. D. and D. T. Campbell (1979) *Quasi Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Co.

Damen, H. and J. W. Delicat (2004) *Samen Leven, Samen Leren: Kansen Voor Jongeren Op Een Zelfstandig Bestaan: Een Evaluatieonderzoek Naar De Foyer De Jeunesse Walcheren*. Nijmegen: Praktikon.

Donner, J. P. H. (2004) *Brief aan de 2e kamer* (Ministerial Letter to the Dutch Parliament), 5299254/04/DJJ. The Hague: Ministry of Justice (17 pp.).

Hirschi, T. (1969) *Causes of Delinquency*. Berkeley, CA: University of California Press.

Hoenderloogroep (1999a) *De Glen Mills School: Beschrijving Van Het Programma. Versie 2*. Twello: Hoenderloogroep.

Hoenderloogroep (1999b) *Glen Mills School: School Voor Winnaars*. Twello: Hoenderloogroep.

Ince, D., M. Beumer, H. Jonkman and M. Vergeer (2004) *Veelbelovend En Effectief: Overzicht Van Preventieve Projecten En Programma's in De Domeinen Gezin, School, Jeugd, Wijk*. Utrecht: NIZW.

Landenberger, N. A. and M. W. Lipsey (2005) 'The Positive Effects of Cognitive-Behavioral Programs for Offenders: A Meta-Analysis of Factors Associated with Effective Treatment', *Journal of Experimental Criminology* 1: 451–76.

Leeman, L. W., J. C. Gibbs and D. Fuller (1993) 'Evaluation of a Multi-Component Group Treatment Program for Delinquents', *Aggressive Behavior* 19: 281–92.

Lipsey, M. W. (1995) 'What do we Learn from 400 Research Studies on the Effectiveness of Treatment with Juvenile Delinquents?', in J. McGuire (ed.) *What Works? Reducing Reoffending*, pp. 63–78. New York: John Wiley.

Lipsey, M. W. (2001) 'Re: Unsolved Problems and Unfinished Business', *American Journal of Evaluation* 22(3): 325–8.

Lipsey, M. W. and D. Wilson (1998) 'Effective Intervention for Serious Juvenile Offenders: A Synthesis of Research', in R. Loeber and D. P. Farrington (eds) *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*, pp. 313–45. Newbury Park, CA: SAGE.

Lipsey, M. W., D. B. Wilson and L. Cothern (2000) 'Effective Intervention for Serious Juvenile Offenders', *Juvenile Justice Bulletin* (April).

Marks, I. M. (2002) 'The Maturing of Therapy: Some Brief Psychotherapies Help Anxiety/Depressive Disorders, But Mechanisms of Action are Unclear', *British Journal of Psychiatry* 180: 200–4.

Michie, S. and C. Abraham (2004) 'Interventions to Change Health Behaviours: Evidence-Based or Evidence-Inspired?', *Psychology and Health* 19(1): 29–49.

Moncher, F. J. and R. J. Prinz (1991) 'Treatment Fidelity in Outcome Studies', *Clinical Psychological Review* 11: 247–66.

Nas, C., D. Brugman and W. Koops (2005) 'Effects of a Multicomponent Peer Intervention Program for Juvenile Delinquents on Moral Judgment, Cognitive Distortions, Social Skills and Recidivism', *Psychology, Crime, and Law* 11: 421–34.

Nigg, C. R., J. P. Allegrante and M. Ory (2002) 'Theory-Comparison and Multiple-Behavior Research: Common Themes Advancing Health Behavior Research', *Health Education Research* 17(5): 670–9.

Polder, W. and F. C. Van Vlaardingen (1992) *Preventiestrategieën in De Praktijk: Een Meta-Evaluatie Van Criminaliteitspreventieprojecten*. Arnhem: Gouda Quint.

Rossi, P. H., H. E. Freeman and M. W. Lipsey (1999) *Evaluation: A Systematical Approach*. Thousand Oaks, CA: SAGE.

Sherman, L. W., D. C. Gotfredson, J. E. MacKenzie, P. Reuter and S. D. Bushway, eds (1997) *Preventing Crime: What Works, What Doesn't, What's Promising?* College Park, MD: Dept of Criminology and Criminal Justice.

Slot, N. W. (1999) 'Competentievergroting Bij Straf, Behandeling En Preventie', paper presented at the Van preventie tot straf, naar meer sociale vaardigheden bij jongeren, de enige weg? (Adolescents: from prevention to punishment to more social abilities, the only way?), 5 November.

Spanjaard, H. (1998) 'Competentievergroting in Een Jeugdinrichting', in W. Koops and N. W. Slot (eds) *Van Lastig Tot Misdadig*, pp. 117–31. Houten: Bohn Stafleu van Lochum.

Sutherland, E. H. (1947) *Principles of Criminology*. Chicago, IL: J. B. Lippincott Co.

Sutton, S. (2003) 'Using Theories of Behaviour Change to Develop and Evaluate Sexual Health Interventions', in J. Stephenson, J. Imrie and C. Bonnell (eds) *Effective Sexual Health Interventions: Issues in Experimental Evaluation*, pp. 51–66. Oxford: Oxford University Press.

Thalheimer, W. and S. Cook (2002) 'How to Calculate Effect Sizes from Published Research Articles: A Simplified Methodology'. URL (consulted March 2004): http://work-learning.com/effect_sizes.htm

Van den Bogaart, P. H. M., K. Mesman Schultz and H. Van der Muijen (2003) *De Glen Mills School: Onderzoeksresultaten Betreffende De Achtergronden Van Studenten En De Behandelingsresultaten*. Groningen: E&M Syntax.

CHRIS BAERVELDT, sociologist, is currently working as a researcher at Utrecht University. His research interests include juvenile delinquency, crime prevention, and social networks. Please address correspondence to: Faculty of Social Sciences, Utrecht University, PO Box 80.140, 3508 TC Utrecht, The Netherlands. [email: c.baerveldt@uu.nl]

BOB HORJUS, educationalist, is working as a researcher and consultant. His research fields include juvenile delinquency, welfare, effect evaluations, and prevention policy. Please address correspondence to: Frank van Borselenlaan 4, 3703 BB Zeist, The Netherlands. [email: horjuspartners@wanadoo.nl]

MICHA DE WINTER is Professor of Social Education at Utrecht University. He is currently working on educational programmes for young criminals and youth participation. Please address correspondence to: Faculty of Social Sciences, Utrecht University, PO Box 80.140, 3508 TC Utrecht, The Netherlands. [email: M.deWinter@uu.nl]