Adult learning; a review of its effects, costs and research

Harvey Goldstein

University of Bristol

1. Introduction

This paper is set out in three sections. The first looks at the papers submitted to the NIACE review of adult/lifelong learning in terms of their contribution to the evidence base. It seeks to draw lessons from these about the problems and insights of existing research. The second section takes up, in detail, one particular approach to the problem of how to allocate resources for adult education. The final section discusses a future research agenda and sets out some proposals for its direction.

2. A review of submissions

The papers reviewed are two by the Matrix knowledge group, two by Andrew Jenkins, one by John Bynner and one by Ricardo Sabates. More detailed critiques are provided in Appendix A. Here we present a summary and general commentary.

The Matrix papers attempt to look at monetary values associated with adult learning, The paper on lifelong learning and crime carries out a cost benefit analysis relating the cost of prison education to monetary benefits in terms of crime reduction. Although it look sat 10 research studies, only one is long term and hence provides credible evidence for recidivism – this issue will be discussed later. One section of the paper attempts to estimate the cost of existing schemes of prisoner education. There is, however, no breakdown by type of education and no attempt to study likely effects of different amounts of spending. The most crucial aspect of the analysis is the computation of social and other benefits associated with a reduction in re-offending, with no distinction made between types of crime. The direct and indirect costs are studied. Direct costs are those associated with the health and judicial system and the indirect ones are those such as social cohesion, suffering etc. There is little discussion of the difficulties of placing monetary values on such things, although the other paper on subjective well-being does make an acknowledgement of these problems and concludes that these difficulties are generally insurmountable. Even for direct costs, however, it is assumed that these are always positive. This, however, is debateable. For example, many murders and violent crimes are gangland ones and the 'elimination' of a gang member, who themselves will have a high probability of offending, is actually a negative cost, saving the cost of both imprisonment and crimes not carried out. I am not condoning the killing of gang members, merely pointing out that the issue is rather more complicated than this paper allows. Indeed, it is not clear that if one carried out a full analysis of all costs that any gain from education would necessarily be positive. Furthermore, it could be argued that one effect of prison education is to reduce the chance of being caught – in which case studies of re-offending convictions miss the point. Again, the Matrix paper does not discuss this. Further comments on cost benefit analysis will be made in the next section.

The other matrix paper looks at the relationship between learning and subjective well being. Using general linear models on the BHPS data they estimate the change in well being for a unit change in the amount of learning. Separately they estimate the change in well being for a unit change in (log) income, From these they equate the amount of learning that is equivalent to a change income - both producing the same change in well being. They point out that we do not understand the factors affecting well-being and that the analyses cannot be interpreted causally. It is difficult to see the precise logic of the argument used. Since the models are limited the results are very sensitive to the variables used in the models and we are given no sensitivity analyses. The paper itself in its discussion admits the weakness of its analysis.

Both these papers exhibit significant technical and conceptual weaknesses and would seem have little to contribute to the commission of inquiry.

The two papers by Jenkins address the issues of adult education and mental health and adult education and children's performance.

The ELSA data are an important resource for studying ageing but they have some important limitations and a major problem with this paper is that these seem not to be recognised nor their effects on inferences assessed.

For example, respondents who were proxies were excluded. These cases are important since they will tend to have a higher proportion of those who became mentally or physically ill and in some cases institutionalised. This will tend to introduce bias and does require discussion. Again, it is not clear why the CES-D scale is dichotomised. This is unnecessary and reduces the sensitivity of the analyses. The control variables include work status and health status. These may well be endogenous since they can be influenced by the 'treatment', namely the acquisition of a qualification between waves 1 & 2. Both these may tend to underestimate any real effects of acquiring a qualification.

There is also an important point in the (unauthored) commentary attached to this paper, in that the amount of depression existing at wave 1 may be important. This suggests that interactions between this variable and the adult learning exposures should be studied.

This paper, because of its technical limitations, presents only weak evidence for its conclusions. A more satisfactory analysis should be done.

The second analysis looks separately at the NPD and LSYPE data. The analysis of the LSYPE is problematic because it uses only wave 1 and is not longitudinal so cannot adjust for prior characteristics. On the other hand the NPD is longitudinal and prior attainment is adjusted for. Thus it makes little sense to compare the results of the two separate analyses and this undermines many of the inferences drawn in this paper. In particular little reliance can be placed on the calculations of changes in attainment to be expected from given changes in neighbourhood deprivation.

As in other papers there are strong assumptions about causality, which are particularly dubious due to the lack of adjustment controls. Overall, therefore, this paper seems to provide scant additional useful evidence.

The Sabates paper looks at the relationship between learning and the reduction of poverty. Overall this paper is disappointing. It attempts an account of what is known through summarising the literature on the relationship between poverty and learning. Unfortunately there is little comment on the quality of the research referred to. Studies in this area are of widely varying quality, as a number of papers have indicated, and what is required is a thorough meta- analytic review that assesses and weights the evidence in terms of its quality and adherence to high research standards. In the absence of this it is not clear that we can place much reliance on the conclusions of this review.

The paper by Bynner looks at the relationship between learning and crime. It provides a welcome exploration of the role of education in crime reduction and prison by arguing for its contextualisation within the judicial, political and general cultural context. It emphasises the need for theoretical perspectives to inform the interpretation and design of studies – especially evaluations. The paper also raises a number of more detailed issues about the conduct of research in this area and provides a basis for further debate. The literature on research in this area is bedevilled by problems of adducing causality, typically from observational (or 'quasi experimental') studies. In an appendix to the paper Bynner discusses the role especially of RCTs and qualitative studies, quoting with approval the stance taken by Campbell that seeks to combine both. Nevertheless, in my view the paper downplays the role of RCTs in evaluation studies. For example, it is not true that these can only deal with one step in a policy process at a time. They can deal with single or composite policies and can extend over time in a sequential fashion. They also are not tied to crude 'hypothesis testing' and should be viewed as part of the evidence accumulation process. Indeed, it is this process of replication of findings in diverse situations that is the key component of scientific theory development and refutation. It should ideally occur using whatever kinds of study are appropriate, and this will generally be a mixture of RCTs, observational and qualitative studies, all of which are greatly enhanced by being longitudinal.

Despite this, the paper does provide a useful and balanced contribution, especially the need to study the kinds of prisoners targetd by education. It also interestingly nakes the point that in the English education system the intense concentration on competitive testing, league tables and targets tends to increase segregation and the extension of this argument into the broader social sphere in terms of the increasing marketisation of services, would be of interest.

There are several general points to emerge from this evaluation of these papers. One is the tentative nature of arguments for causal relationships. The existing research is relatively weak in this area and this implies that any estimates of costs and benefits should be treated with caution. Another issue that emerges is that the research appears to concentrate on the already educationally disadvantaged, yet there will be important evidence available from those not so disadvantages who participate in lifelong learning. A third issue is the need to disaggregate education and its recipients into the different types of delivery and people. This will be important for the targeting of policy.

As with the other papers reviewed, the emphasis appears to be on the study of the effects of education on the already educationally disadvantaged. Yet, the life course history of criminals and prisoners who are not educationally disadvantaged is also relevant and contains information that could be valuable both for dealing with their own lives as well as those disadvantaged. The reasons those with relatively high educational achievements enter crime would be illuminating to study. Fourthly, the general assumption made in these papers is that the cost of prison is the money spent on the prison service on a per head value. Yet the real reduction in cost is the marginal cost for the prisoners who would not re-offend as a result of education (assuming we can estimate these numbers) since the basic costs of maintaining a prison service will remain. Presumably the marginal cost is smaller, perhaps greatly so, in which case the financial case is weakened.

A fifth issue is the need to embed analyses of adult learning within the wider socio-political context. Actions by governments and others will affect learning opportunities and perceptions as well as rewards and need to be incorporated in analyses as far as possible.

Finally, another issue that emerges from these papers and from the research they quote is the overall low quality of the quantitative research. This is an area that needs serious attention, with much greater concentration on establishing good practice.

3. Modelling the allocation of resources for adult learning

This section will address the issue of 'investment' in lifelong learning and the 'return' that might be expected from such investment. The terms 'investment' and 'return' are used in a very broad sense to include, for the former, any resource that is devoted to the particular (lifelong learning) activity being considered, and for the latter any measurable outcome associated with the activity. We shall elaborate further in subsequent sections.

The traditional approach to such an issue is to try to quantify, in monetary terms, the costs of the investment and any subsequent return. This typically takes to form of Cost-Benefit Analysis (CBA) and one of the principal challenges facing this approach is to find ways of reducing both investments and returns to a single monetary scale. Since, generally, there is no objective method for doing this, analysts will use sensitivity strategies to study the robustness of assumptions and public debate often centres on ways of quantifying benefits in monetary terms.

A further, related, issue is that returns are rarely unidimensional, and perceived benefits in one area may be associated with losses in other areas. The classic example is where the development of a new drug prolongs life (a benefit) but is associated with increased costs of prolonged medical care (a loss). In the context of lifelong learning, a return in terms of enhanced workplace skills for a group of workers would generally be perceived as a benefit, but might also have a loss associated with a shortage of new workers to undertake the (key) jobs that were previously done by that group of workers. This is an example of the need to embed any models within the context in which they are expected to operate, and to accept also that the parameters of those models, the values that will determine how they perform, may change with the particular context in which they are

applied. We shall argue later that this is an important motivation for taking seriously the setting up of research projects in diverse circumstances with a view to a comparative evaluation of their achievements.

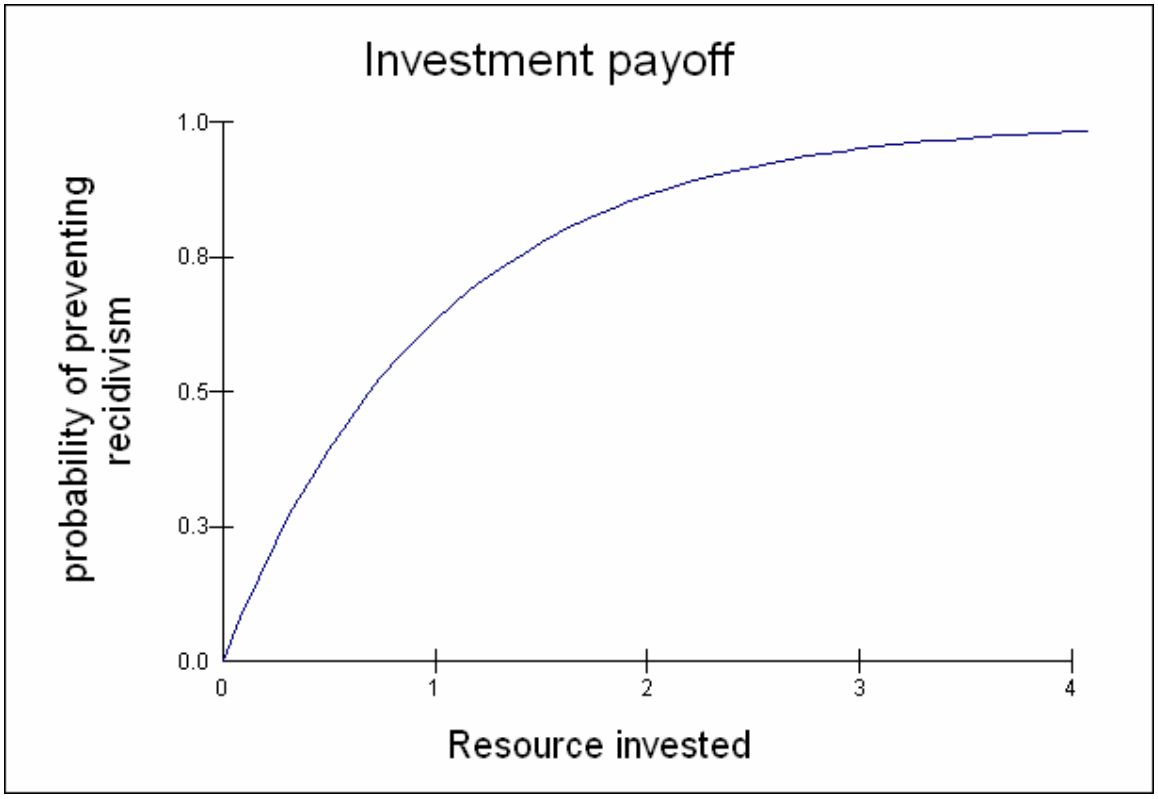
In the next section we introduce an approach to studying returns to investments that avoids the 'input-output comparability' problems associated with CBA by considering how a *given amount* of investment can be allocated optimally among members of a population.

3. Optimal allocation of resources

To introduce our model we consider the example of prison education aimed at reducing recidivism.

We begin by assuming that a given fixed amount of resource is available and we shall consider principally just one kind of resource, let us say that a certain number of teachers, n_t, is available. These teachers are to be allocated among prisons and prisoners, and the default allocation is assumed to be a random allocation where each teacher is assigned for a basic unit of time (say one day) to a prisoner chosen at random. We shall consider alternative allocation schemes, but first we look at the relationship between the exposure to a teacher for different numbers of days and the probability of preventing recidivism *for a prisoner who would otherwise become a recidivist*. A very general form for such a relationship is shown in Figure 1.

Figure 1. Relationship between resources allocated (arbitrary scale) and probability of preventing recidivism.



The basic assumption we are making is that the more resources we allocate the greater the probability that we will prevent recidivism, and that beyond a certain point 'diminishing returns' set in, in the sense that we need increasing amounts of the resources to raise the probability by any given amount. The zero point on this scale is arbitrary since, presumably, some resources will always be available, so that we shall assume that the resources we are concerned with allocating are those additional to anything that already exists, and we assume for simplicity that if no additional resources are used there is a zero probability of preventing recidivism, that is a randomly chosen prisoner will become a recidivist. If we wished, however, we could include existing resources in our measure of resources, thus relocating the zero point. In a general sense a relationship such as that in Figure 1is bound to occur, but what we will need to know in any particular case is the exact nature of the relationship, since this will be crucial for our model. Alberman and Goldstein (1970) and Goldstein (1972) consider this in more detail, and provide a general justification for the 'negative exponential' curve which is depicted in Figure 1.

So far we have assumed implicitly that each prisoner will respond in the same way to a given amount of resource. In practice, however, prisoners will have different existing propensities to recidivise and we would expect therefore, that the response to a given resource will depend on that underlying propensity. These propensities, expressed as probabilities, can range from 0 to 1. We shall therefore also assume that we can classify prisoners by this propensity, or alternatively derive an estimate of this probability, based upon prisoner characteristics and other variables such as prison characteristics.

2.1 A simple example

To make matters more concrete consider a simple numerical example. Let us suppose that we have a prisoner with a propensity 0.5 to recidivise. We can think of this as the prisoner recidivising on just half the occasions (given no resources) in a notional repetition of their sentence. Suppose also that at each repetition an amount of resource *r* is used and this results in a probability of avoiding recidivism of q_r, i.e. recidivism does not happen in a proportion q_r of occasions when otherwise the prisoner would recidivise. For the 50% of occasions when the prisoner does not recidivise, the allocation of resource *r* clearly does not change the outcome. Thus, out of all the occasions the proportion of occasions where recidivism does not occur given resources *r* is just $q_r/2$. More generally we can see that if a prisoner has propensity *p* to recidivise and if a given resource, *r*, is associated with a probability q_r of avoiding recidivism, for a prisoner who would otherwise become a recidivist, then the probability of avoiding recidivism is simply pq_r.

We now look, in an informal way, at the consequences of different resource allocations, in terms of recidivism rates. Suppose that we have available a total amount of resource, *RN*, where N is the number of prisoners. We also suppose that we can classify prisoners into two groups, a low risk group with a basic recidivism rate of 30% and a high risk group with a basic recidivism rate of 80%, and that the latter group comprises just 10% of the prison population. Suppose that our resources are fairly limited so that allocating an amount R to a 'recidivist' prisoner (one who would recidivise given no resource) is associated with a 30% probability of not recidivising – this corresponds in Figure 1 to a resource value of (about) 0.4. First, consider allocating all the available resources equally

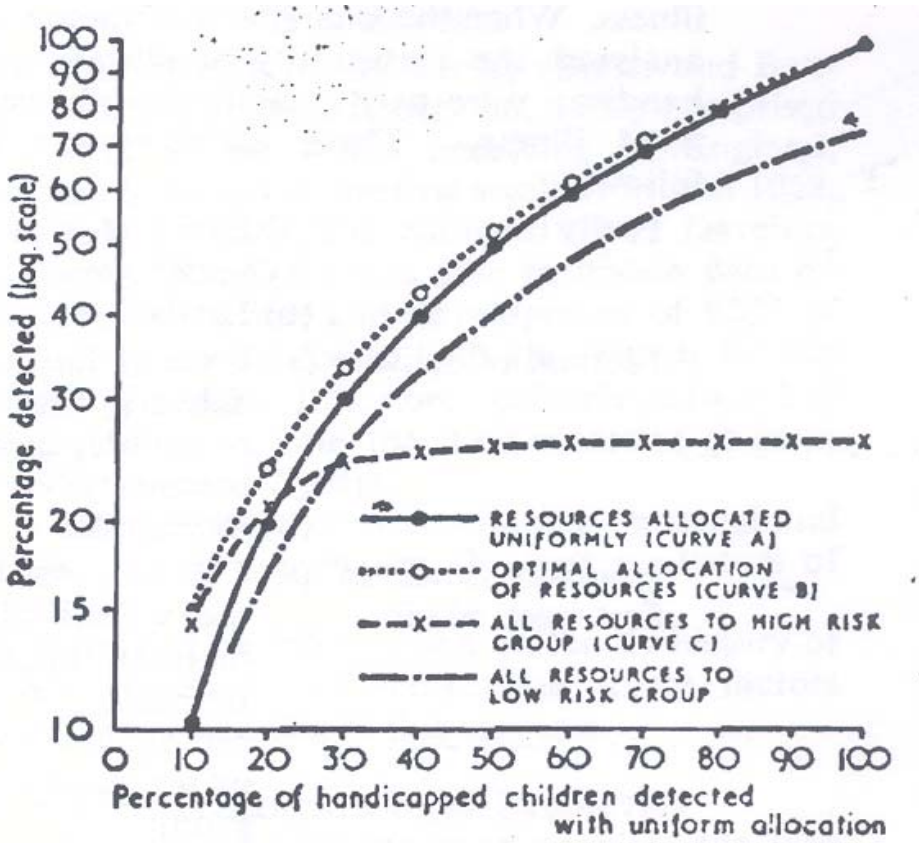
to every prisoner in the high risk group – a strategy that might often seem an appealing way of using available resources. This implies that each member of this group obtains a resource 10*R*. If we refer to Figure 1 and a resource value of 10 x 0.4 = 4 we see that the probability of not recidivising is (say) 0.96. Thus, for the 0.80N/10 prisoners in the high risk group who would otherwise recidivise only (1-0.96) x 0.80 x N/10 = 0.0032N actually do so. Among the low risk group the number who recidivise is just 0.30 x 9N/10=0.27N, so that overall the number who recidivise is 0.273N, i.e. just 27.3%.

Now consider allocating all the resources equally to every prisoner in the low risk group. We can now go through an analogous set of calculations and we find that we now allocate 10*R*/9 resources to each prisoner – corresponding, let us say, to a 35% probability of not recidivising. Thus for the 0.30 x 9N/10 who would otherwise recidivise only 0.65 x 0.30 x 9N/10 = 0.175N actually do so and the number who recidivise in the high risk group is now 0.80N/10=0.08N and the total number is therefore 0.251N or just 25.1% We see therefore that the apparently counter-intuitive allocation of all resources to the low risk group actually results in a lower recidivism rate.

The reason, of course, is that in the high risk group the maximum possible number that we can prevent from recidivising is just the number who are recidivists, that is 0.1N x 0.8=0.08N (our first resource allocation actually prevents 0.068N), whereas our second resource allocation to the low risk group actually prevents 0.095N from recidivising; there simply are more recidivisms to be prevented in the low risk group, and however many resources we allocate to the high risk group, we can never prevent more than 0.080N and this is fewer than the we can prevent by giving all resources to the low risk group. It is clear therefore that giving everything to the high risk group is, beyond a certain point, counterproductive with diminishing returns. In fact there is an optimum strategy which is intermediate between the two we have illustrated, and it will involve giving *relatively* more to each member of the high risk group than to members of the low risk group. The actual relative amounts can be calculated if we have the required information about the relationship, such as in Figure 1.

We show this with a real life example from another application. Goldstein and Alberman (1970) applied this resource allocation model to the detection of handicaps in children, using a classification into two risk categories based upon data from the National Child Development Study. The aim was to maximise the number of handicaps detected using screening resources applied according to the risk. In Figure 2 the results of using different strategies is illustrated. This shows clearly that, when resources are relatively scarce an allocation solely to the highest risk group is indeed optimum, whereas with plenty of resources such a policy becomes less desirable and finally very counterproductive, and worse than giving all to the low risk group. There is, however, always an optimum allocation, just as we have argued in our example. The resources are measured on a scale in terms of the percentage of handicaps detected given a uniform allocation to all children. A similar scale definition would be convenient for our prison population, and would require that a given amount of resources is measured in terms of the proportion of recidivists it prevents form recidivising when everybody receives the same amount. To establish what this would be, of course, requires empirical data and this issue will be returned to later.

Figure 2. Resource allocation strategies for a childhood screening programme. From Alberman and Goldstein (1970). High risk group contains 13.2% of children.



The full statistical details are given by Alberman and Goldstein (1970) and Goldstein (1972) extends the model to consider the case where there is more than one potential outcome and more than one type of screening resource available, but we shall here only consider the basic model of one resource type and one outcome. We note, however, that we are not limited to just two risk groups, and in general the more differentiated the population is in terms of risk the better our resource allocation procedure will be.

4. Further examples

To see how our resource allocation modelling might apply in other cases we shall look at three examples. The first is the prevention of depression among students with no qualifications. We assume that this can be achieved by enabling students to achieve a 'level 2' qualification by allocating resources to assist them. In this case we would seek to predict depression and divide the students into risk groups. This prediction might use family characteristics or contextual ones such as the nature of their learning environment. Given an estimate of the relationship between resources allocated and probability of avoiding depression, as in Figure 1, our resource allocation procedure would follow along the lines discussed above.

Our second example concerns investment to improve health literacy among GP patients. Here, our outcome is a reduction in the proportion of patients who do not understand how to use drugs they are prescribed. Patients may be classified into risk groups, for example based upon educational background, and a knowledge of the Figure 1 relationship may be determined from experiments with different amounts and types of resources (see below).

Our third example concerns the prevention of elderly people losing their independence. Again, dividing into risk groups can be done given available research data and knowledge of the Figure 1 relationship obtained empirically.

4. Cost implications

There are several aspects of costs involved with the procedures discussed. One of these is concerned with the differential administration of resources. Another is concerned with the costs of obtaining the data needed for the modelling, and a third is concerned with monitoring and evaluating any system. These costs need to be studied, and it is conceivable that they may be large enough to rule out the use of the procedures in some cases. We shall not, however, consider this further, save to remark that in all the examples considered above, any programme will have its attendant costs and these can be compared directly in monetary terms. The costs of monitoring any given policy is generally not negligible and should be considered at the outset, and we shall discuss monitoring more fully below.

5. Measuring resources and predicting risk

We have already indicated how a ‘resource scale’ can be set up. The crucial information required is knowledge of the relationship between the resource allocated and the response in terms of preventing an event occurring. It is not difficult generally to see how one might do this in principle. In the recidivism example one might experimentally assign at random different amounts of educational resource to different groups and observe the resulting recidivism rates. Apart from possible ethical issues, such a study would need to be long term and involve rather large numbers and hence be quite expensive – but still possible. An alternative might be to use existing data on recidivism rates matched to whatever resource exposure had been measured for individual prisoners, and to carry out an analysis of the relationship, attempting to adjust for confounding factors etc. In fact, even an approximate estimate of the relationship will generally be useful in guiding resource allocation, and this is one of the areas for more detailed research (see below).

For the prediction of risk, there will often be existing data that can be used. We note that we are not here principally interested in causal relationships, merely in ways of efficiently predicting a risk, for example of recidivism, in order to group individuals into resource reception groups. Among the factors that will typically need to be taken into account, especially for adult learning, is that of the institution, so that data analysis should use some form of multilevel modelling that is able to do this.

6. A research agenda

There are two key aspects to a future research agenda. One concerns ways of obtaining information about the allocation-prevention relationship and the other concerns the evaluation of any schemes that are set up using the models described. These are of course related, at least in the sense that the results of the latter can be expected to inform the former. A further research issue is a technical one that relates to the nature of the models themselves and this will be addressed briefly.

We have already seen that the optimum resource allocation depends on knowledge of the allocation-prevention relationship. We have also seen, however, that useful general statements about allocation, such as not allocating everything to a high risk group when resources are plentiful, can be made when only the general *form* of the relationship is known. We would expect, therefore that even an approximate knowledge of the relationship will be useful and it may be possible to obtain this from existing data where resources have been differentially allocated and records allow matching to outcomes. If a resource allocation programme is set up then the study of the allocation-prevention relationship should be incorporated. In practice, any such programme will use resources of different types and it should be designed so that the separate, and interactive, effects of each resource can be evaluated. In general, it will also be possible to cost each type of resource so that the monetary cost of each possible combination of resources can be used to form a common resource scale (see also Goldstein, 1972 for a discussion).

The importance of evaluating a resource allocation programme cannot be overestimated: it constitutes the only sure way of knowing whether a programme is a success. Any such programme should be designed so that different procedures are used, and so that the model can be tested and refined, possibly as a continuous process. There will typically be ethical constraints on how much resource provision can be varied, and debate about these will be needed, but such requirements would not, as far as can be envisaged, undermine the possibility of this kind of resource allocation research. As already mentioned an important aspect of such evaluations is to study 'side effects' or drawbacks and to attempt to quantify these. Strictly speaking such effects lie outside the consideration of the models described, but that does not diminish the importance of measuring them.

There are features of the current model as set out in this paper that may need elaboration before the model can be implemented widely. We have already mentioned the need to consider several types of resource and different outcomes; these are discussed by Goldstein (1972). In particular, the notion of different outcomes can be extended to outcomes that are incidental effects on other people with whom the individual has contact with. In the recidivism example this might for example be families of prisoners.

A further practical issue is that an outcome may not be binary – either an event such as recidivism occurs or it doesn't. We may wish to consider degrees of recidivism or the time to recidivism as a factor, so that *levels* of outcome are involved. One way of dealing with this is to assign 'values' or 'scores' to the different levels so that these are maximised (or minimised) and this would be a useful theoretical extension. Another issue is that the current model assumes each individual responds independently of any other individual. This may not always be the case, especially in situations of learning where individuals are grouped together and will receive a common resource, such as a teacher.

It may well be the case that group dynamics affect the outcome and this should therefore be taken into account in the resource allocation model, and also in any evaluation. Likewise the costs associated with a resource will typically depend on how it is administered, in a group or individual setting. Thus, for example, an educational resource in a prison may be more cheaply allocated to a given number of prisoners by choosing them all from the same prison rather than from several prisons.

Finally, we should emphasise again the manner in which the approach of this paper is different from traditional CBA in avoiding the need to place all costs and benefits on a single monetary scale. Of course, it will always be necessary to decide how many resources can be made available, although very often this is fixed more or less by the availability of manpower or physical facilities. Even where it is not, consideration of the optimal allocation of different overall amounts should provide useful insights.

References

Alberman, E. D. and H. Goldstein (1970). "The at risk register: A statistical evaluation." British Journal of Social and Preventive Medicine **24**: 129-135.

Brooks, G., Burton, M., Cole, P., Miles, J., Torgerson, C., Torgerson, D. (2008). Randomised controlled trials of incentives to improve attendance at adult literacy classes. Oxford Review of Education, 34, 493-504.

Goldstein, H. (1972). "The allocation of resources in population screening: a decision theory model." Biometrics **28**: 499-518.

McGuire, J. (2008). A review of effective interventions for reducing aggression and violence. Philosophical Transactions of the Royal Society, 363(1503), 2483-2622.

Postscript

Since completing the main report, and following further discussions[1] and literature review, some additional suggestions for future research directions have emerged. In this not I shall look first at the nature of the criterion used for judging the success of prison education and by extension to other forms of adult education, and secondly at some aspects of study design.

Is recidivism an appropriate criterion for judging the success of prison education programmes?

In prison there are typically many diverse programmes and 'treatments' available. These include those that can be described as educational, often at a basic level, as well as those that can be described as behavioural, especially those dealing with extreme, addictive or other behaviours related to the offence committed. A feature of this latter group as opposed to educational programmes is that they are *directly* aimed at reducing criminal behaviour, so that recidivism is the appropriate outcome measure which a particular treatment or programme is designed to influence. There is a considerable literature devoted to trials of this kind (see e.g. McGuire, 2008 for a recent meta-analysis). Education programmes, on the other hand, generally are not specifically designed to reduce recidivism, and as pointed out in the main report, it can be argued that a conceivable outcome is that they will reduce the likelihood that recidivism is observed, without necessarily altering the rate, and so create a spurious positive effect.

Thus, in terms of evaluating, and hence possibly justifying prison education, it would seem both misguided and possibly counterproductive to attempt to justify it in terms of what at best could be considered a side-effect of any programme. Rather, the success of any education programme should be judged in *educational* terms. In the next section I shall outline some technicalities of such evaluations, but it is worth pointing out that these observations better set into context the somewhat equivocal evidence about the relationship between prison education and recidivism. It may also be the case that currently the apparent emphasis on studying this relationship may owe more to a desire to persuade policymakers' that educational programmes directly address their concerns with reducing crime, than to actually demonstrating that educational levels of prisoners have improved. If this is indeed the motivation within the adult education community then it would seem to be seriously misplaced and ultimately counterproductive.

Evaluating educational programmes

Given what has been said about the various programmes to which prisoners are exposed, the evaluation of any of these needs to take account of all of these potential influences. I shall consider two contexts; one is the use of some kind of randomised trial and the other is the analysis of data obtained during the course of a prisoner's incarceration.

---

[1] I am most indebted to Professor Stephen Wong for several illuminating discussions about research with prisoners.

Randomised trials

To carry out a randomised trial of an educational (or any other) programme where randomisation takes place within a prison, is beset with difficulties. The administrative difficulties are generally severe since prison routines are difficult to manipulate. Even if contamination between treatment arms can be avoided for prisoners, doing so for staff will be more difficult. Ensuring cooperation of prisoners can be problematic, and differential dropout may occur. For all these reasons it would be better to consider cluster randomised trials where whole prison units are assigned to different treatments. In general this will imply large numbers of participating prisons. Although such studies may be expensive, if they are confined to evaluating educational outcomes as suggested above, they may be less expensive than follow up studies of recidivism.

If successfully carried out a randomised trial may indicate that certain programmes are relatively successful. In addition, however, it will be important to understand why these may perform better and in particular what it is about the prison context that may interact with particular programmes or features of programmes to achieve success. Thus, information about the prison regime, and about the characteristics of the prisoners will be important to study, especially any other programmes or activities that they are involved with.

The study of interactions is of particular interest. Many existing adult education evaluations have been designed to detect *average* effects (see for example Brooks et al., 2008 who used just 14 treatment and 14 control classes) but what is often of greater interest is to know whether effects vary for particular groups. This implies much larger sample sizes than typically used. Thus, for example, the fact that a study finds very small average effects may mask important large interactions for certain groups with important implications for policy. This issue is relevant to non-randomised studies also.

Observational studies of prisoners

In the absence of the possibility of carrying out a randomised trial, information can still be collected on prisoner characteristics, including the nature of their offences and their backgrounds, as well as characteristics of other programmes in operation and the prison regime. Both qualitative and quantitative information will be useful.

While this postscript has discussed prison education, many of the issues apply more generally. In particular two are worth emphasising. First, is the need to concentrate on those outcomes that educational programmes are directly meant to address, rather than more remote outcomes such as economic benefits. Secondly is the need to take account of other associated activities experienced by participants. These issues apply to randomised trials as well as non-randomised ‘observational’ studies.