Jnl Soc. Pol., 34, 2, 175–194 © 2005 Cambridge University Press
doi:10.1017/S0047279404008530 Printed in the United Kingdom

The Perilous Road from Evidence to Policy: Five Journeys Compared

ANNETTE BOAZ* and RAY PAWSON**

**Queen Mary, University of London*
email: a.l.boaz@qmul.ac.uk
***Contact author: Department of Sociology and Social Policy, University of Leeds, Leeds LS2 9JT*
email r.d.pawson@leeds.ac.uk

Abstract

Comprehensive reviews of the available research are generally considered to be the cornerstone of contemporary efforts to establish 'evidence-based policy'. This article provides an examination of the potential of this stratagem, using the case study of 'mentoring' programmes. Mentoring initiatives (and allied schemes such as 'coaching', 'counselling', 'peer education' and so on) are to be found in every corner of public policy. Researchers have been no less energetic, producing a huge body of evidence on the process and outcomes of such interventions. Reviewers, accordingly, have plenty to get their teeth into and, by now, there are numerous reports offering review-based advice on the benefits of mentoring. The article asks whether the sum total of these efforts, as represented by five contemporary reviews, is a useful tool for guiding policy and practice. Our analysis is a cause for some pessimism. We note a propensity for delivering unequivocal policy verdicts on the basis of ambiguous evidence. Even more disconcertingly, the five reviews head off on different judgemental tangents, one set of recommendations appearing to gainsay the next. The article refrains from recommending the ejection of evidence baby and policy bathwater but suggests that much closer attention needs to be paid to the explanatory scope of systematic reviews.

Introduction

The apparatus of evidence-based policy and practice is well established. The 'systematic review' of the available evidence has emerged as the favoured instrument, and research syntheses are now commissioned and conducted right across the policy waterfront. The advantages of going beyond single studies to appraise the body of knowledge relevant to a given policy or practice question are palpable. The case for systematic review is put most famously in Lipsey's compelling metaphor: 'What can you build with thousands of bricks?' (Lipsey, 1997). His answer is that it is high time to put aside solitary evaluations, which tend to come up with answers that range from the quick-and-dirty to the overdue-and-ambivalent. These can and should be replaced with the considered appraisal

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

of the collective findings of dozens, hundreds and, just occasionally, thousands of primary studies, so constructing a solid citadel of evidence.

Not all commentators have been so optimistic about the value and contribution of research reviews. Bero and Jadad (1997) suggest that, while reviews have an obvious appeal as an objective summary of a large quantity of research, there is little evidence to suggest that they are actually used to inform policy and practice. Kitson *et al.* (1998) argue, moreover, that the debate about evidence-based policy and practice focuses on the level and nature of the evidence at the expense of an understanding of the environment in which the review is to be used and the manner in which research is translated into practice. They suggest that there has been an implicit assumption that it is only the nature of the evidence that is of real significance in promoting good-quality, usable reviews.

This tension is put to empirical scrutiny in this article, using materials gathered as part of a larger study being conducted within the current UK ESRC Research Methods Programme.[1] The policy intervention under inspection is 'mentoring'. Whether the 'kindness of strangers' is an appropriate cornerstone for all social reform is a moot point (Freedman, 1999), but mentoring initiatives and their cousins such as 'coaching', 'counselling' and 'peer education' are to be found in every corner of public policy. There are indeed a thousand (and more) evidential bricks to be assembled, thus making mentoring an ideal test bed of the potential of evidence-based policy.

The sheer weight of such evidence has inspired over a score of research teams to conduct reviews of the available research on mentoring and, of these, five are considered here. This article does not seek to pass technical judgement on their conduct. They utilise quite different strategies for synthesising evidence, but there is no attempt to award methodological gold stars or wooden spoons. Instead the reports are examined in respect of the advice offered to the policy maker. What is the nature of the recommendations in each review? How might the policy maker choose between them? Do they generate proposals with genuine policy import and, importantly, do the five syntheses speak with one voice?

It transpires that many different viewpoints flow from the reviews. Indeed, there is a whole range of incompatibilities and, at their heart, some seemingly contradictory advice on whether mentoring can be recommended for at-risk youth. If this is a typical picture, and there is every reason to believe it is, serious consequences flow for the endeavour of evidence-based policy. There is already enough discord in the ranks of policy pundits without it being replicated by the evidence underlabourers.

The article moves on to discuss reasons for the mixed messages. Some of the discrepancies in counsel originate from the variation in methods used in synthesising evidence. Others flow from subtle differences in the questions posed (or commissioned) in each review. And yet others flow from inconsistency in the selection and coverage of the primary studies included in the synthesis. We

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

take some pains to describe these disparities in the gestation of our reviews before coming to our central contention. As well as compressing huge bodies of evidence, reviews labour under the expectation that the synthesis is for policy's sake. Thus, over and above all their technical differences and under the self-imposed pressure to deliver clear policy recommendations, there is a tendency to inflate the conclusions. The article will show in some detail that, at the point of making recommendations, there is an inclination to 'go beyond the evidence'.

There is powerful weaponry here for those observers who suppose that social science can never get beyond methodological debate, paradigm wars and petty rivalries. Does not such a Babel of briefings allow policy makers to pick and choose between the reviews, and thus change the nature of the beast to policy-based evidence? We do not in fact share this gloomy conclusion. The underlying reason for the mixed messages on mentoring is the sheer complexity of evaluative questions that need to be addressed. One can review bygone evidence not only to ask whether a type of intervention 'works' but also in relation to 'for whom, in what circumstances, in what respects and why it might work'. For good measure, a review might also be sensibly aimed at quite different policy and practice questions such as how an intervention is best implemented, whether it is cost-effective and whether it might join up or jar with other existing provision.

The evidential bricks can be cemented together in a multitude of edifices and thus only modest, conditional and focused advice should be expected from research synthesis. The reputation of systematic reviews has suffered badly from the foolhardy claims of early advocates who argued that they would deliver pass/fail verdicts on whole families of initiatives (Sherman *et al.*, 1997). And while the five reviews discussed here are much less 'gung-ho', we still detect signs of over-ambition in the search for *res judicata*.

In our conclusions we turn to a potential solution to this problem. The truth of the matter is that reviews are non-definitive. Painstaking and comprehensive they may be, but the last word they are not. Once we are rid of the notion that there is a gold-standard method of research synthesis capable of providing unambiguous verdicts on programmes, and once we jettison the notion that a single review can deliver all-purpose policy advice, there is a way forward. What have to be developed are portfolios of reviews and research, each element aimed at making a contribution to the explanatory whole.

Five reviews

We sought out existing reviews on 'mentoring programmes' in order to explore the ways in which they constitute an evidence base to support policy and practice. Within this sample, syntheses of 'youth mentoring' were in the highest concentration and so we jettisoned reviews on, for instance, mentoring for nursing and teachers from our study (Andrews and Wallis, 1999; Wang and

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

Odell, 2002). Of the studies on youth mentoring, five were selected that were conducted in recent years, that provided sufficient detail on their own analytic strategy and that explicitly described themselves as having a review function. We were particularly interested in comparing approaches that spanned the spectrum of approaches to review and synthesis.[2] Those selected range in approach from a formal 'meta-analysis' of randomised controlled trials to a narrative 'literature review' and an 'evidence nugget' designed to be of direct use to practitioners and decision makers. The rationale here, to repeat, is not to make a direct methodological comparison but to investigate rather different hypotheses about the digestibility of the rival strategies into policy making and their respective potential to embrace or resist the intrusion of the reviewers' own policy preferences.

Table 1 lists the reviews, profiles the basic strategy and (in bold) notes the authors' description of their review activity. We also provide an initial summary of the key findings in the third column.

Our title contemplates 'a journey' from evidence to policy. We put it like this in recognition of countless studies of research utilisation, which have shown there is no simple, linear progression from research report to programme implementation (Hogwood, 2001; Lavis *et al.*, 2002). We thus begin our exploration of the utility of our fist of reviews with a 'thought experiment', comparing them in terms of their respective capacities to penetrate highways and byways of policy making. How might a decision maker decide which of these reviews to use? Suppose they landed on the desk an official with responsibility for mentoring: which might strike a chord, which might be considered the most informative? We recognise, of course, yet further simplifications assumed in these ruminations of our imaginary bureaucrat. In reality, whole ranks of policy makers and practitioners and committee structures have to be traversed (Schwartz and Rosen, 2004). Nevertheless, we hope to indicate just of few of the very many reasons, apart from quality of evidence, that may help or hinder utilisation of these very real research products.

As a small but not entirely incidental aside here, we note that in some of our discussions with government researchers and officials with policy responsibilities in this area we suggested this as a real exercise. The assembled mound of review material was edged on to various desks and we offered to leave it behind for closer scrutiny. The response was – how shall we put it – not one of overwhelming gratitude. It would seem that, as utility road-block one, evidence is welcome only in somewhat more pre-digested and definitive chunks.

This reluctance at the water's edge notwithstanding, let us proceed by picturing our chimerical official wading through the reviews. Some of the later reviews make reference to their predecessors, but we begin by considering them one at a time before contemplating the composite, if distinctly fragmented, picture.

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

Review 1

Our policy maker might be tempted to go with the evidence of effectiveness presented by DuBois *et al.* (2002) who offer a clear conclusion (that mentoring 'works', albeit with moderate to small effects), and a list of helpful 'moderators' indicating some of the conditions (such as 'high-risk' youth) and best practices (such as closer monitoring), which are shown to enhance the overall effect. This meta-analysis has the quality stamp of publication in a peer-reviewed journal and draws on leading expertise in both mentoring (DuBois) and the chosen methodology (Valentine and Cooper). However, it is questionable how far the reader would stray past the best practices and bottom-line conclusion and into the 40 pages of dense methodological description. Few policy makers, we suppose, would be quick with an opinion on whether the authors were correct in 'windsorising' outliers in coming to their calculation of net effects.

Does this matter? Within a medical context there is an assumption that review users will be in a position to carry out a basic critical appraisal in order to decide whether or not to use the findings to shape policy and practice. Training courses are run by the Institute of Child Health[3] and the Critical Appraisal Skills Programme (CASP)[4] to develop these skills in practitioners. Such an assumption is questionable within social policy fields where consensus is lacking on what counts as good evidence, and on how best to synthesise it. In the probable absence of shared methodological wisdom, one has to contemplate whether the policy maker might prefer to place emphasis on the clarity and persuasiveness of abstracts, summaries and conclusions, as well as on quality filters such as peer review that are imposed by others.

Review 2

Roberts's 'phenomenological' review (Roberts, 2000), we suppose, would trigger a different reaction in the policy community. The first issue is to ponder whether it would be regarded as 'evidence' at all? The review begins with a discourse on social science epistemology, in which the perspectives of Wittgenstein and Husserl are called upon to justify the idea that reviews should 'clarify' rather than seek to judge 'what works' (so setting up a tension with Reviews 1, 3, 4 and 5). On the other hand, it is hard to deny that this review provides the most compelling and comprehensive picture of all the components, contours and complexities of mentoring interventions. It distills an overall model that is likely to resonate with those policy makers and practitioners concerned with the details of programme implementation. Whether they would share Roberts's exact vision of mentoring is a more debatable point, however, as is his suggestion that he has reached and refined his model on the basis of identifying 'consensus' in the literature.

Given its odd mixture of exposition and assertion, and given that 'presuppositionless phenomenological reduction' is not the driving heartbeat

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

TABLE 1. Five reviews: a summary.

Author and title	Type of review	Main findings
Review 1		
DuBois D, Holloway B, Valentine J & Cooper H (2002) *Effectiveness of mentoring programs for youth: a meta-analytic review*	A 40-page '**meta-analysis**' published as a journal article. The paper is written by US academics, drawing on US experimental evaluations of mentoring programmes. Only studies with comparison groups are included. The review draws on literature from 1970–1998 and aims to assess the overall effects of mentoring programmes on youth and investigate possible variation in programme impact related to key aspects of programme design and implementation.	There is evidence that mentoring programmes are effective interventions (although the effect is relatively small). For mentoring programmes to be as successful as possible, programmes need to follow guidelines for effective practice. Programme characteristics that appear to make a difference in promoting effective practice include on-going training for mentors, structured activities, frequent contact between mentors and mentees and parental involvement.
Review 2		
Roberts A (2000) *Mentoring revisited: a phenomenological reading of the literature*	A 26-page '**phenomenological review**' published as a journal article. The paper is written by a UK academic and mentor, drawing on a wide range of literature from a long time period. The review aims to contribute to our understanding of what we mean by the term 'mentoring.'	There is consensus in the literature that mentoring has the following essential characteristics: it is a process, a supportive relationship, a helping process, a teaching–learning process, a reflective process, a career development process, a formalised process and a role constructed by or for a mentor. Coaching, sponsoring, role modelling, assessing and informal process are deemed to be contingent characteristics.
Review 3		
Lucas P & Liabo K (2003) *One-to-one, non-directive mentoring programmes have not been shown to improve behaviour in young people involved in offending or anti-social activities*	A 13-page '**evidence nugget**' published as a web report. This report is written by UK academics, using reviews and primary studies from the UK and other countries. The review draws on literature from 1992 to 2003 focusing on mentoring with young people who are involved in offending or other types of anti-social activities. The review focuses on the impact of mentoring on outcomes for young people, but also examines the resource implications and alternative strategies to promote behaviour change.	Mentoring programmes have not been shown to be effective with young people who are already truanting, involved in criminal activities, misusing substances or who are aggressive. There is evidence to suggest that mentoring programmes might even have negative impacts on young people exhibiting personal vulnerabilities. However, mentoring may have a preventative impact on young people who have not yet engaged in anti-social activities.

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

Review 4		
Hall J (2002) *Mentoring and young people: a literature review*	A 45-page '**literature review**' commissioned by a government department and conducted by a UK academic. The review includes a range of study types to address the different questions within the scope of the review. It also draws on existing reviews. The review focuses on research published between 1995 and 2002 and concentrates on mentoring to support 16–24 year olds accessing and using education, training and employment. The review addresses a range of questions including 'what is mentoring', 'does it work' and 'what are the experiences of mentors and mentees?'	Mentoring is an ill-defined and contested concept. The US evidence suggests that mentoring is an effective intervention, although the impact may be small. Successful mentoring schemes are likely to include: programme monitoring, screening of mentors, matching, training, supervision, structured activities, parental support and involvement, frequent contact and ongoing relationships. The UK literature concludes that mentoring needs to be integrated with other activities, interventions and organisational contexts. Most mentors are female, white and middle class, and report positive personal outcomes including increased self-esteem.
Review 5		
Jekielek S, More K A & Hair E C (2002a) *Mentoring programs and youth development: a synthesis*	A 60-page '**synthesis**' (with a stand alone eight-page summary version) conducted by an independent, not-for-profit US research centre for an American foundation working with low-income communities. The review includes different study types to address different questions within the review, and draws on both primary studies and review evidence. The review includes literature from 1975 to 2000 and focuses on the role of mentoring in youth development. The review looks at a range of questions including 'what do mentoring programmes look like', 'how do they contribute to youth deve lopment' and 'what are the characteristics of successful mentoring'?	Mentoring has a positive impact on youth development in terms of education and cognitive attainment, health and safety, and social and emotional well-being. Positive outcomes for youth participating in mentoring programmes include: fewer unexcused absences from school, better attitudes and behaviours at school, better chance of attending college, less drug and alcohol use, improved relationships with parents and more positive attitudes towards elders and helping others. Programme characteristics associated with positive youth outcomes include: mentoring relationships that last more than 12 months, frequent contact between mentor and mentee and youth-centred mentor–mentee relationships. Short-lived mentoring relationships have the potential to harm young people, cross-race matches are as successful as same-race matches and mentees who are the most disadvantaged or at-risk are especially likely to gain from mentoring programmes.

of Whitehall, and given that the review is buried in a specialist and relatively low-status corner of the academic literature, a question mark has to be raised on whether this study would make it to the evidence-based policy starting blocks.

Review 3

The third review is produced with the policy and practice communities clearly in its sights, and the decision maker might well be attracted to the brief and clearly presented 'Evidence Nugget' (Lucas and Liabo, 2003). The research team does not claim to produce a full-scale systematic review, and readers are relieved of the need to thumb through a telephone directory of appraisals of primary studies. The synthesis is made on the back of an examination of an existing review (Review 1 above), a methodological critique of some of the best-known primary studies on the Big Brother/Big Sister programmes, and findings from a small selection of other evaluations, including UK programmes. Despite a significant overlap in source materials, the authors reach a much less positive conclusion than do the other four syntheses. Such programmes 'cannot be recommended' as an intervention of proven effectiveness for young people with personal vulnerabilities and with severe behavioural problems. Indeed, the authors cite evidence that harmful effects can follow mentoring for such troubled youth, and advise policy makers to look elsewhere for more effective interventions in such cases.

Should the policy maker have as much trust in this review? Clearly, it is much more selective than most traditional reviews, and the reason for choosing this rather than another admixture of reviews, appraisals and case studies is not made clear. It is a 'web-only' product and as such might not carry the formal weight of a peer-reviewed publication. This format, however, does allow for update and revision, and the authors have carried out their own peer review (using a named panel) in amending an original draft. Finally, in terms of the 'provenance' of the piece, the authors and their group are well known in the fields of review methodology, and policy and practice interventions for young people.

Review 4

The Hall review (2002) was commissioned by a specific policy agency (the Scottish Executive's Enterprise and Lifelong Learning Department) and, as such, an emphasis on accessibility to the decision maker is to be expected. The report is printed in hard copy and is also freely available on the web. The language is accessible and the exposition painstaking, with recommendations carried in executive, sectional and sub-sectional summaries. The main contrast with the other reviews is in terms of ground coverage, with the author attempting to answer a wide range of questions including: what is mentoring, does it work, what makes it work, how is it viewed by different stakeholders, and should it be regulated? The overall tone is one of neither enthusiasm nor opposition.

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

Sometimes Hall simply concludes that research has little to say on certain policy issues. In respect of the crucial 'does it work?' issue, he sides with the findings and indeed the technical wisdom of Review 1.

For our policy reader, there may be some questions about the quality of this report. Its tasks are so many and varied that there might be doubt about the veracity of all the conclusions. Questions might be raised about the lack of attention to context, in terms of the applicability of much of the reviewed material to the Scottish population and polity. More generally, the typical suspicions that surround 'literature' reviews could be raised. There is little concrete exposition of the methodology employed, or mention of expertise in the fields of mentoring or reviewing. There is also no clear indication of how, or indeed whether, the report was appraised by peers prior to publication.

Review 5

Jekielek *et al*. (2002a) offer a full report and an eight-page summary (2002b) of their review on mentoring strategies. It concentrates on outcomes, as do Reviews 1 and 3, but does so at much lower levels of aggregation. That is to say, it examines intervention effects on a huge number of attitudes and behaviour (school attendance, drug use, relationships with parents and so on). It also reviews research on the implementation characteristics of effective mentoring programmes (frequency of contact, cross-race matching, level of risk of mentee and so on). The main body of the research carries these findings in a score and more tables. For each and every potential outcome change (such as high-school grades), a table enumerates and identifies the primary studies, tallies relative successes and failures, and lists some of the key programme characteristics that are associated with the more successful outcomes. The appendix provides ultimate disaggregation in the form of a glossary of each original study. There is, however, a short and clearly written summary (Jekielek, 2002b) offering an analysis of the implications for policy and practice.

Though it has none of the technical complexities of Review 1, the multifaceted and highly conditional findings may encourage policy-oriented readers to leap directly to the summary. The report does use clear quality criteria in the selection of primary studies (though it admits some well-known studies found wanting by Review 3). The synthesis is clearly well funded, of some status, and accessible (published, via the web, jointly by Child Trends and the Edna McConnell Clark Foundation). As a postscript it might also be noted that, like Review 1, it draws its evidence entirely from the US, and UK users might feel concern about relying closely on the analysis.

Evidence or opinion?

So where do we stand? Remember that these are a mere selection of the available reviews. The evidence has been sorted and appraised, and we appear to have a

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

range of rocky outcrops rather than the hoped-for brick tower. What we have shown is that evidence never comes forth in some 'pure' form; it is always mediated in production and presentation, and that various characteristics of the chosen medium might well be significant in establishing the policy message. It is far from clear how a decision maker might use these reviews to help inform policy and practice questions.

Such a conundrum assumes, of course, that the decision maker is seeking the best-quality evidence to help resolve an open question. For the policy maker hoping to lever resources into mentoring initiatives, a positive synthesis (Review 5) about a 'promising strategy' can be selected judiciously from this pile. For the policy maker preferring the firing squad, the Evidence Nugget (Review 3) provides handy ammunition. At a level down from the cynical choice, we have the pragmatic choice. For the decision-maker in a hurry, selection might be narrowed down to the short summary version of the evidence base produced with the practitioner in mind (Review 3 or the 'briefing' of Review 5). For the government analyst seeking a general overview of the topic, Hall's work (Review 4) may well appeal, with its broad coverage of relevant issues and questions. Finally, of course, it should be acknowledged that it might not be the report itself that proves influential. For example, commissioning loyalties, coverage of the research in the press, a presentation from the authors, or the use of the research by a lobbying organisation may have the vital impact on the decision maker.

But what of the rational choice? Do the rival reviews merely replicate the conceptual disunity and paradigm wars that are all too familiar in mainstream social science? Is there a way of making sense of the diversity of recommendations that apparently reside in these five studies? One way to choose between them is to subject them to further methodological scrutiny and to appraise them, proposition by proposition, on the grounds of technical rigour. We are unsure, however, whether dragging the debate one step backward towards research, epistemological and ontological fundamentals would (a) end in agreement or (b) help with policy or practice decisions. We did not imbue our imaginary decision maker with a great deal of patience, but we are reasonably sure that the majority of (real) policy makers will not be particularly interested in, or equipped to make, such methodological judgements.

So, our approach here is to look for a more general malaise that might underlie the disorder. Our claim is that a false expectation, going by the name of the 'quest for certainty', has gripped those conducting research synthesis. It is this collective ambition to generate concrete propositions that decision makers can 'set store by' that is the root of the problem. The working assumption is that the review will somehow unmask the truth and shed direct light on a tangible policy decision. In the summative passages of the typical systematic review, the evidence becomes the policy decision. The consequence is that when 'evidence-based' recommendations are propelled forth into the policy community, they

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

often shed the qualifications and scope conditions that follow from the way synthesis has been achieved.

We believe that over-ambition of this sort infects reviews of all types, and many types of inferences that reside within them. Here we identify two typical ways in which reviews over-extend themselves and illustrate the point with several examples from our case studies. While we believe we have uncovered a general shortcoming, it should be noted that the argument does not apply with perfect uniformity to all the studies. Nor, of course, are the following points meant to offer a comprehensive analysis of each review. To repeat for emphasis, we are interested in the moment of crystallisation of the policy advice. Our aim is to show that, at this point of inference, doubt has the habit of being cast into certainty. And in the following analysis we seek to make a contrast between a list of (numbered) policy recommendations and the body of evidence from which they are drawn.[5]

Seeing shadows, surmising solids

First, we concentrate on the most dangerous of all questions on which to aspire to certitude: namely, does it work? This is the question that meta-analysis is designed to answer and readers should note the rather modest results in this genre produced by Review 1 (median effect size, $d = 1.8$). On the basis of this analysis DuBois *et al.* assert:

> From an applied perspective, findings offer support for the continued implementation of and dissemination of mentoring programmes for youth. The strongest empirical basis exists for utilising mentoring as a preventative intervention for youth whose backgrounds include significant conditions of environmental risk and disadvantage. (Statement 1)

The auspices here are quite good enough for Hall (Review 4) in the section of his review dealing with the efficacy of youth mentoring. Dubois *et al.*'s conclusions are quoted at great length, and verbatim, on the basis that:

> This is a highly technical, statistically based analysis with a strong quantitative base which has been conducted entirely independently of any of the mentoring schemes reviewed. As such it must be given a great deal of weight. (Statement 2)

Despite its fine-grained portrayal of outcome variations, more unequivocal policy pronouncements are to be found in Jekielek *et al.*'s research brief (Review 5).

> The most important policy implication that emerges from our review of rigorous experimental evaluations of mentoring programmes is that these programs appear to be worth the investment. The finding that highly disadvantaged youth may benefit the most reinforces this point. (Statement 3)

What can been seen here is the foregathering of 'definitive' statements. They positively beckon the policy maker's highlighter pen. Note further that these

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

viewpoints, once pronounced, have a habit of becoming ensconced as authentic evidence in subsequent literature. Take another statement from Jekielek *et al.*:

> Mentored youth are likely to have fewer absences from school, better attitudes towards school, fewer incidents of hitting others, less drug and alcohol abuse, more positive attitudes to their elders and helping in general, and improved relationships with their parents. (Statement 4)

Again, this is reproduced by Hall in Review 4, albeit with slightly more caution on the basis that, compared to DuBois *et al.*, Jekielek *et al.*'s report is 'less extensive . . . and reported in less detail'.

So far so good for youth mentoring. The evidence, or perhaps the rhetorical use of evidence, seems to be piling up. But now we come to the first jarring contradiction. The Evidence Nugget (Review 3) concludes:

> On the evidence to date, mentoring programmes do not appear to be a promising intervention for young people who are currently at risk of permanent school exclusion, those with very poor school attendance, those involved in criminal behaviour, those with histories of aggressive behaviour, and those already involved with welfare agencies. (Statement 5)

How can this be? Not only is there disagreement on the overall efficacy of youth mentoring, the 'at risk' group – singled out previously as the prime focus of success – is now highlighted as the point of failure. This is strange indeed because the reviews all call upon a similar body of evidence, with the Evidence Nugget making use of its two predecessors.

Let us examine first the summative verdict. Dubois *et al.*'s net effect calculations revealed a 'small but significant' effect on the average youth. But by the time it reaches policy recommendations, this datum becomes transmogrified by the original authors into significant-enough-to-continue-implementation and then, by contrast, into small-enough-to-look-elsewhere by Lucas and Liabo. The suggestion of the Evidence Nugget team is that:

> In view of the research evidence, it may be prudent to consider alternative interventions where larger behavioural changes have been demonstrated such as some form of parent training and cognitive behavioural therapy. (Statement 6)

This conclusion is reached, as per our thesis, without reference to any supporting evidence on the two alternative interventions. In our view there is very little mileage in blanket declarations about wholesale programmes. In actuality, the evidence as presented cannot decide between the above two inferences on the overall efficacy of mentoring. They are matters of judgement. They are decisions about whether the glass is half full or half empty. Our point, nevertheless, is that the authors seem content or perhaps compelled to make them, and make them, moreover, in the name of evidence.

Can the difference of opinion on the utility of mentoring for 'at risk' youngsters be explained and reconciled? Again, we perceive that the difficulty lies with the gap between the pluck of the pronouncement and the murk of

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

the evidence. Jekielek *et al.* (Review 5) explain this against-the-odds success by way of a typical 'pattern'. That is to say, Sponsor-a-Scholar youth with the least parental and school support who entered schemes with low initial GPAs advanced more than did those with good prior achievement who tended to 'remain on the plateau'. Other programmes are said to follow along 'similar' lines.

Meta-analysis pools together the outcomes of very many programmes, successful and unsuccessful, in coming to an overall verdict. It is also able to investigate some of the factors, known as 'moderators' and 'mediators', which might generate these different outcomes. On this basis, DuBois *et al.* are able to say a little bit more about the identity of the high-risk group.

> Effect sizes were largest for samples of youth experiencing both individual and environmental risk factors or environmental risk factors alone. Average effect sizes were somewhat lower for the relatively small number of samples in which youth were not experiencing either type of risk. (Statement 7)

So far so good for success with the at-risk group that policy interventions have found so hard to reach. But then we have to imagine our policy maker coming across the long and jarring line of the untouchable high-risk categories claimed in Statement 5. This disappointing conclusion is reinforced further in the 'practice recommendations' of the Evidence Nugget, as follows:

> Caution should be used when recommending an intervention with a group of youngsters at raised risk of adverse outcomes . . . There is evidence for example that peer group support for young people with anti-social behaviours (which falls beyond the scope of this Evidence Nugget) may exacerbate anti-social behaviour, increasing criminal behaviour, antisocial behaviour and unemployment in both the short and the long term. (Statement 8)

So who is right? The conclusions on risk of Reviews 1 and 5 stand in stark contrast to many qualitative studies of youth mentoring (see, for example, Colley, 2003). These show that the disadvantaged and dispossessed are very unlikely to get anywhere near a mentoring programme and that when they are compelled to do so the relationship comes under severe strain. It is also hard to square with many process evaluations of mentoring (Rhodes, 2002) that show considerable pre-programme drop out among 'hard-to-reach' mentees as they face frequent long delays before a mentor becomes available. Jekielek *et al.*'s Research Brief (2002b) also provides an interesting caveat:

> Some very at-risk young people did not make it into the Sponsor-A-Scholar program. To be eligible, youth had to show evidence of motivation and had to be free of problems that would tax the program beyond its capabilities.

The very positive statements (1, 3 and 7) thus actually refer to a rather curious sub-section of those who can be considered 'high risk'. They appear to emanate from the worst socio-economic backgrounds and have suffered high degrees of personal trauma, but consist only of a subset who have had the foresight

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

to volunteer for a mentoring programme, and the forbearance to wait for an opportunity to join it.

So should the policy maker take the lead from the Evidence Nugget? The evidence raised on the potential perils of mentoring high-risk youth is limited to quite specific encounters. Reference is made to a study that discovered declines in self-esteem following broken and short-term mentoring partnerships. Youth in such relationships tended to have been 'referred for psychological or educational programs or had sustained emotional, sexual or physical abuse'. The other negative instance (see Statement 8) is decidedly peripheral and comes from studies of 'peer education' programmes. All sorts of quite distinctive issues are raised in such schemes about whether one lot of peers can reverse the influence of a different lot of peers.

Our conclusion is that there are some terribly fine lines, yet to be drawn in and thus yet to be extracted from the literature, on which type of youth, with which backgrounds and experiences will benefit from mentoring. Jekielek *et al.*'s review includes rather commonplace matters such as the role of educational underachievement in conferring risk. DuBois *et al.* use an (undefined) distinction between 'environmental' and 'personal' conditions as the source of risk. When operationalised, it leaves only a relatively small (but undisclosed) number of mentees who are not at risk (note this detail in Statement 7). Lucas and Liabo spell out a more specific set of conditions to identify heightened vulnerability (Statement 5), although it is not clear that their negative evidence corresponds to each of these risk categories.

The concept of 'risk' carries totemic significance in all policy advice about young people, but remains shrouded in mystery. We submit, in this instance, that the opinions on risk of all three review teams are likely to carry a hint of truth. But by adopting a slightly different take on how to describe the vulnerabilities that confront young people and on how to identify them in the source materials, it is possible to deliver sharply different conclusions. The problem, as noted above, is that fuzzy inferences are then dressed and delivered as hard evidence. They are deemed to speak for the 'evidence base'; they are celebrated as being 'highly technical' and for the 'strength of the statistical analysis'; they are endowed with the solidity of 'nuggets'.

Studying apples, talking fruit

Since its invention, systematic review has struggled with the problem of 'comparing like with like'. In a critique of the earliest meta-analysis of the efficacy of psychotherapeutic programmes, Gallo (1978) opined that actual interventions were so dissimilar that 'apples and oranges' had been brought together in the calculation of the mean effect. Here we highlight a somewhat different version of the same problem. The issue, once again, is the leap from the materials covered in the evidence base to the way they are described when it come to proffering policy

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

and practice advice. What typically happens is that the review will (perforce) examine a subset of research on a subset of trials of a particular intervention, but will then drift into recommendations that cover the entire family of such programmes. One review will look at apples, the next at oranges, but the policy talk is of fruit. This disparity crops up in our selection of reviews, both explicitly (in terms of attempts to come to a formal definition of mentoring) and implicitly (in the *de facto* selection of certain types of mentoring programmes for review).

It is appropriate to begin with Roberts (Review 2), who sets himself the task of distilling the phenomenological essence of mentoring. He attempts to 'cut through the quagmire' by distinguishing mentoring's essential attributes from those that should be considered contingent (see Table 1 for the key features). This cuts no ice with Hall (Review 4), who also reviews mentoring terminology and comes to the opposite conclusion, namely that mentoring is an ill-defined and essentially contested concept, and that the wise reviewer will pay heed to the diversity of mentoring forms rather than closing on a preference. Lucas and Liabo's (Review 3) more pessimistic conclusions feature 'non-directive' mentoring programmes. Their definition covers some of the same ground as Roberts and Hall, although there is stress on those programmes in which the mentor is a volunteer, and on the delivery mechanisms which are about support, understanding, experience and advice.

The other two reviews take a more pragmatic approach to defining mentoring. They concede that there are differences in 'goals, emphasis and structure' (Review 5) and in 'recruiting, training and supervision' (Review 1). The precise anatomy of the mentoring programmes under inspection is therefore defined operationally. Jekielek *et al.* review a selection of named programmes identified by their umbrella organisations (Big Brothers/Big Sisters, Sponsor-A-Scholar and so on). DuBois *et al.*'s operational focus is established by the search terms (such as 'mentor', 'Big Brother') used to identify the relevant programmes, and by the inclusion criteria (for example, 'before-after and control comparisons') used to select studies for review.

We capture here a glimpse of the bane of research synthesis. Do the primary studies define the chosen policy instrument in the same way *and* do reviewers select for analysis those that follow their preferred definition and substantiate their policy advice? In respect of our five studies of mentoring, once again we see them heading off into slightly different and rather ill-distinguished territories. We do not claim to have captured all the subtle differences in definitional scope in the brief remarks above. Nor, to repeat for emphasis, are we claiming that there is a correct and incorrect usage of the term 'mentoring'. What is clear is that some subtle and not-so-subtle, and some intentional and not-so-intentional, differences have come into play.

Recall, that our purpose is to focus on the potential user's understanding of the reviews. In this respect we make two points. The first is that these

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

terminological contortions are often rather well hidden and disconnected from the advice that emanates from the reports. Recommendations are often couched in rather generic terms, the report titles (see Table 1) referring to 'mentoring programmes for youth', 'mentoring and young people' and so on. The reader can also be usefully referred back to the list of key *statements* extracted above. Instead of reading them for claims about whether the programmes work or not, they can also be inspected for slippage into rather broad references to the initiatives under review, with the use of non-specific terms such as 'mentoring programmes', 'mentored youth' and so on.

The second problem concerns the issue of generalisation and the transferability of findings. For policy makers and practitioners, a basic concern is what happens on 'their patch' and thus what *type* of mentoring programmes might make headway and which should sensibly be avoided for their particular clients. In this respect, the definitional diversity across the various reviews leaves decision makers with some tricky inferential leaps. If they want to follow the successes identified in the Jekielek *et al*. review, they are advised to go for a 'developmental' as opposed to a 'prescriptive' approach. If they want to avoid the perils of mentoring as identified by Lucas and Liabo, they are directed to a promising example based on 'directive' as opposed to 'non-directive' techniques. Once again, we sound to be heading towards contradiction. The former preference appears to cover items like 'frequent contact', 'flexibility' and 'mentee-centered' approaches. The latter is described in terms of 'frequent contact', 'advocacy' and 'behavioural contracting'. It might be, therefore, that there is more overlap than is suggested in the bald advice. But our critical point remains. In order to take this fragment of evidence forward, the user would have to proceed on the basis of guess-work rather than ground-work.

Conclusion: refocusing reviews

Systematic reviews have come to the fore partly as a response to an over-reliance on 'expert consultants', whose advice is always open to the charge of cronyism. Our five examples, however, demonstrate the 'non-definitive' nature of reviews: even where they focus on similar questions, they can come to subtly (sometimes wildly) different conclusions. What is more, certain reviews are unlikely to make it to the policy forum. Roberts's review (Review 2) was produced as a personal quest within an academic context, and it is unlikely that such an effort would ever be commissioned formally.

So is this all a presage of the fate of the evidence-based policy movement? Are we bound to end up with squabbling reviews and overlooked evidence? And in the all-too-probable absence of the methodological power to adjudicate between contending claims, will decision makers in search of advice be forced to fall back on reputation? Will they end up in the arms of another kind of 'expert', one who has punch in the systematic review paradigm wars?

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

In fact, we draw a more positive conclusion from our review of reviews. Looking at these five attempts to synthesise the mentoring literature, we have argued that reviewers committed to providing evidence for policy and practice have focused their energies on methods rather than utilisation, and on the quest for certainty rather than explanation. While we do not suppose for a moment that the techniques of selecting and systematising findings are unimportant, we argue that this leaves unexplored a bigger, strategic question about the overall usage of evidence. And that issue is the need for clearer identification of the policy or practice questions to which the evidence is asked to speak.

We have tried to demonstrate that no single review can provide a definitive 'answer' to support a policy decision. One review will not fit all. Why this sample of reviews struggles and why contradictions emerge is because of the lurking emphasis on the 'what works' question. But as soon as this mighty issue is interrogated, it begins to break down into another set of imponderables. It is clear that mentoring can take a variety of forms, so we need to review the working mechanisms: what is it about different forms of mentoring that produces change? It is also clear that mentoring has a rather complex footprint of successes and failures, so we need to review the contextual boundaries: for whom, in what circumstances, in what respects, and at what costs are changes brought about? The answer to the efficacy question is made up of resolutions to all of the tiny process and positioning issues that occur on the way to the goal.

In another part of our project, we have engaged in discussions with policy makers and government researchers with responsibility for mentoring interventions. They have identified a veritable shopping list of questions and issues that are crucial to mentoring *and* that are worthy of review. These include much more specific questions about effectiveness (for example, does matching mentors to mentees affect outcomes, does mentoring need to be buttressed with other forms of welfare support), and detailed questions relating to implementation (for example, how important is the training and accreditation of mentors, should mentors be volunteers or be paid, how important is pre-selection and so on).

We thus reach our main conclusions on the need for reviews that respond in a more focused way to such a shopping list of decision points. Does this mean that we favour the more comprehensive approach to evidence synthesis, adopted in Hall's intrepidly wide-ranging review (Review 4)? At the risk of apparent contradiction (and of seeming forever prickly) our answer is not entirely positive. Such multi-purpose reviews tend to exhaust themselves, with the result that they are often broad in scope but thin on detailed analysis. Let us raise the very brief example of one of Hall's tasks, namely to synthesise the available material on the mentees' perspectives. In the midst of all his other objectives, Hall simply grinds to a halt: 'there is little literature that explores the views of mentees in any depth'. Having embarked on a rather similar mission, we can only disagree. Research

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

synthesis is a mind-numbing task, relevant information is squirreled away in all corners of the literature, and it may be more sensible to go one step at a time.

Our way forward is to match the exploration of the evidence base to the complexity of the policy decisions. In the case of mentoring programmes, this will require a 'portfolio' of reviews to explore the numerous questions of interest to policy makers and practitioners that we have begun to enumerate above. The vision is of the deployment of a half-a-dozen or so reviews, each with a clear division of labour in terms of the analytic question/decision point under synthesis. And on this point, at least, we can bring our five reviews onside. At first sight it is hard to get past their inconsistencies and contradictions, but if one looks to the compatibilities and commonalties, a rather more engaging possibility presents itself. A collective picture emerges from this work that mentoring is no universal panacea and that it should be targeted at rather different individuals in rather different circumstances. The composite identikit of the ideal client seems to be of a tergivasating individual who is at once highly troubled and highly motivated to seek help. Imagine what would happen if we targeted a review to solve this precise conundrum. Imagine if there were a handful of other reviews orchestrated to answer other compelling questions. Imagine, further, that these reviews were dovetailed with ongoing, developmental evaluations. We might then be in a position to talk about evidence-based policy.

On a final (and more sober!) note, we return to the policy maker at the water's edge. Before getting as far as reconciling perplexing and contradictory evidence, many policy makers (and their analytical support staff) appear to be put off by the style and format of reviews. As a next stage of this project we have conducted interviews with government researchers and policy makers to explore the ways in which they could be tempted to take the plunge and have a closer look at the evidence. So there is at least one further desideratum for our portfolio of reviews: they need not only to be cumulative and technically proficient, but attuned to policy and practice purposes *and* presented in attractive and usable ways.

Acknowledgement

The authors would like to thank (in a manner of speaking!) the editors and two anonymous reviewers for a compendious list of astute suggestions on clarifying the arguments made herein. The remaining errors and simplifications remain our own work.

Notes

1 http://www.ccsr.ac.uk/methods/.

2 We have attempted to use neutral terminology of 'reviews', 'syntheses', 'overviews' and so on in covering and crossing from one approach to the other. Sometimes we use the term 'systematic review' but, again, that is not meant to bestow methodological privilege on any review so described, for we agree with Hammersley (2001) that all methods of synthesis

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

employ tacit as well as pre-formulated strategies. Our sample of review styles also inevitably omits some of the developing approaches to research synthesis for the simple reason that they have yet to be deployed on youth mentoring. But a further comparison along the lines conducted here can eventually be performed in relation to meta-ethnography, realist synthesis, Bayesian meta-analysis, the EPPI approach and so on. Background material on all of these can be found in Dixon-Woods *et al*. (2004) and at www.evidencenetwork.org.

3 http://www.ich.ucl.ac.uk/ich/html/academicunits/paed_epid/cebch/about.html.

4 http://www.phru.org.uk/~casp/casp.htm.

5 A referee has raised a taxing question on our strategy here, one which in fact challenges the entire 'act of compression' that always occurs in conducting a review. Put simply, our argument is that, at certain points in the reviews under study, the authors are less cautious in their policy advice than is warranted by their own stockpiles of evidence. Unavoidably, we make that claim on the basis of a selective presentation of the policy assertions and a highly compressed account of the review strategies and findings. Ipso facto, could it not be that we too are being selective in presenting those fragments of the original reviews that suit our own case? Our answer is that the highlighted policy pronouncements are produced verbatim and at sufficient length to confirm that they do indeed set forth on a favoured policy agendum. In particular, our case is made in the demonstration that the authors favour different policy conclusions on the basis of similar primary materials. More crucial than this, however, is the fact that our thesis and investigatory tracks are made clear enough so that they can be checked out and challenged. It is open to anyone, including the original authors, to deny the inferences drawn. And it is open to us, if challenged, to supply further instances of overstatement from the same body of materials. Our constant refrain throughout this article is that reviews and, perforce, reviews of reviews do not have a methodologically privileged position and are thus never definitive (for example, see Marchant's, 2004, critique of a recent Campbell review). Trustworthy reviews stem from organised distrust. Much more could be said on the basic philosophy underlying this view of objectivity, but it may be of interest to report that it is a version of what Donald Campbell himself calls 'competitive cross-validation' (Campbell and Russo, 1999).

References

Andrews, M. and Wallis, M. (1999), 'Mentorship in nursing: a literature review', *Journal of Advanced Nursing*, 29: 1, 201–7.

Bero, L. and Jadad, A. (1997), 'How consumers and policymakers can use systematic reviews for decision making', *Annals of Internal Medicine*, 127: 1, 37–42.

Campbell, D. and Russo, M. (1999), *Social Experimentation*, Thousand Oaks: Sage.

Colley, H. (2003), *Mentoring for Social Inclusion: A Critical Approach to Nurturing Mentor Relationships*, London: Routledge Farmer.

Dixon-Woods, M., Agarwal, S., Jones, D. and Sutton, A. (2004), 'Integrative approaches to qualitative and quantitative evidence', UK Health Development Agency, Available at: www.hda.nhs.uk/documents/integrative_approaches.pdf.

DuBois, D., Holloway, B., Valentine, J. and Cooper, H. (2002), 'Effectiveness of mentoring programs for youth: a meta-analytic review', *American Journal of Community Psychology*, 30: 2, 157–97.

Freedman, M. (1999), *The Kindness of Strangers: Adult Mentors, Urban Youth and the New Voluntarism*, Cambridge: Cambridge University Press.

Gallo, P. (1978), 'Meta analysis: a mixed metaphor?', *American Psychologist*, 33: 5, 515–17.

Hall, J. (2002), 'Mentoring and young people: a literature review', The SCRE Centre, 61 Dublin Street, Edinburgh EH3 6NL, 67 pp. (Research Report 114). Available at: http://www.scre.ac.uk/resreport/pdf/114.pdf.

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press

Hammersley, M. (2001), 'On "systematic" reviews of research literatures: a "narrative" response to Evans and Benfield', *British Education Research Journal*, 27: 5, 543–54.

Hogwood, B. (2001), 'Beyond muddling through – can analysis assist in designing policies that deliver?', in *Modern Policy-Making: Ensuring Policies Deliver Value for Money*, National Audit Office publication Appendix 1. Available at: http://www.nao.gov.uk/publications/nao_reports/01-02/0102289app.pdf.

Jekielek, S., More, K. and Hair, E. (2002a), 'Mentoring programs and youth development: a synthesis', Child Trends Inc, 430 Connecticut Avenue NW, Suite 100, Washington DC 20008, 68 pp. Available at: http://www.childtrends.org/PDF/MentoringSynthesisfinal2.6.02Jan.pdf.

Jekielek, S., More, K., Hair, E. and Scarupa, H. (2002b), 'Mentoring: a promising strategy for youth development', Child Trends Inc, 430 Connecticut Avenue NW, Suite 100, Washington DC 20008, 8 pp. (Research Brief). Available at: http://www.childtrends.org/PDF/mentoringbrief2002.pdf.

Kitson, A., Harvey, G. and McCormack, B. (1998), 'Enabling the implementation of evidence based practice: a conceptual framework', *Quality in Healthcare*, 7: 3, 49–158.

Lavis, J., Ross, S. and Hurley, J. (2002), 'Examining the role of health services research on public policymaking', *The Milbank Quarterly*, 80: 1, 125–54.

Lipsey, M. (1997), 'What can you build with thousands of bricks? Musings on the cumulation of knowledge in program evaluation', in D. Rog and D. Fournier (eds.), *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods*, New Directions for Evaluation 76, San Francisco: Jossey Bass, pp. 7–24.

Lucas, P. and Liabo, K. (2003), 'One-to-one, non-directive mentoring programmes have not been shown to improve behaviour in young people involved in offending or anti-social activities', What Works for Children, 14 pp. (Evidence Nugget). Available at: http://www.whatworksforchildren.org.uk/nugget_summaries.htm.

Marchant, P. (2004), 'A demonstration that the claim that brighter lighting reduces crime is unfounded', *British Journal of Criminology*, 44: 3, 441–7.

Rhodes, J. (2002), *Stand By Me: The Risks and Rewards of Mentoring Today's Youth*, Cambridge, MA: Harvard University Press.

Roberts, A. (2000), 'Mentoring revisited: a phenomenological reading of the literature', *Mentoring and Tutoring*, 8: 2, 145–70.

Schwartz, R. and Rosen, B. (2004), 'The politics of evidence-based health policy-making', *Public Money and Management*, 24: 2, 121–7.

Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P. and Shawn, D. (1997), 'Preventing crime: what works, what doesn't, and what's promising', US Department of Justice, 810 Seventh Street NW, DC 20531, 483 pp. Available at: http://www.ncjrs.org/works/wholedoc.htm.

Wang, J. and Odell, S. (2002), 'Mentored learning to teach according to standards-based reform: A critical review', *Review of Educational Research*, 72: 3, 481–546.

https://doi.org/10.1017/S0047279404008530 Published online by Cambridge University Press