Making the Most of the Evidence: Evidence-based policy in the classroom

Nick Cowen, Kings College London and
Nancy Cartwright, Durham University with
Baljinder Virk and Stella Mascarenhas-Keyes

CHESS Working Paper No. 2015-03
Durham University
July 2015

CENTRE FOR HUMANITIES
CHESS
ENGAGING SCIENCE AND SOCIETY

CHESS working paper (Online) ISSN 2053-2660

Making the Most of the Evidence: Evidence-based policy in the classroom

Nick Cowen and Nancy Cartwright with
Baljinder Virk and Stella Mascarenhas-Keyes

Nick Cowen
Department of Political Economy
Kings College London
Second Floor
Strand Building
Strand Campus
London WC2R 2LS
Email: nicholas.cowen <at> kcl.ac.uk

Nancy Cartwright
Philosophy Department
Durham University
50 Old Elvet
Durham DH1 3HN
Email: nancy.cartwright <at> durham.ac.uk

Abstract: What allows research evidence to contribute to successful social policy and improve practice in public services? The establishment of the What Works Network, a group of evidence 'clearing houses', that summarise academic research evidence for practitioners in different policy areas, heralds a resurgence of evidence-based approaches in British policymaking. This report outlines the existing literature analysing the process of implementing evidence-informed policy, with a particular focus on the field of primary and secondary education. New data, based on interviews with teachers in primary and secondary schools, along with the analysis of existing literature, suggest that understanding the rationale for a particular policy approach and recognising relevant contextual factors are crucial for selecting and implementing successful policies. This suggests that local knowledge, as well as professional judgement and expertise, are critical contributors to policymaking alongside formal scientific research evidence.

Research methods. This report is based on a 3-month study supported by a secondment of Nick Cowen (Department of Political Economy, King's College London) to the Cabinet Office, working with Professor Nancy Cartwright of Durham University and University of California. They were supported by Dr Baljinder Virk (National Audit Office) and Dr Stella Mascarenhas-Keyes (Department for Education) who were seconded for part-time roles using the Government Social Research Network.

We survey and critically analyse academic discussions of evidence-based policy implementation published in the last decade and a half, looking for central issues and identifying common themes. This literature review is not systematic and is based on informal searches of the scholarly literature on implementing evidence-based policy and some analysis of influential policy documents. For primary data collection, we adopted an 'elite interview' method, interviewing 22 teachers at 12 school sites in a variety of English geographic locations and a range of different student characteristics. They were identified using snowball sampling techniques and approaches on social media.

All views expressed are those of the authors and not of any mentioned organisations or government departments.

1. Executive Summary

1.1 Existing literature

- Theoretical accounts of implementing evidence-based policy focus on medicine and public health as paradigm fields. Strong support for evidence-based policy approaches is present, as well as radical opposition, frequently from critical theory and interpretivist perspectives. Some criticisms identify challenges of causal inference, especially when interpreting the results of randomised controlled trials. Crucially, an average of observed outcomes of an intervention does not produce a reliable indicator of what that intervention will produce in a particular context.
- Empirical research on the implementation of evidence-based policy shows that 'evidence' itself is defined in a variety of different ways across different institutional environments and, for practitioners, does not always refer to academic research or scientific methods. In education contexts, local knowledge was identified as crucial for successful implementation. Some approaches identify fidelity to the evidence-based approach as important, while others emphasise the importance of adapting approaches to local circumstances.
- The Education Endowment Foundation's Teaching and Learning Toolkit represents a unique contribution to the existing literature. It systematically summarises research evidence in a way that is aimed at engaging education professionals. This seems to be unmatched in the rest of the evidence-based literature on education.

1.2 Results of this study

- This study is based on interviews with 22 teachers at 12 school sites in a variety of English geographic locations and a range of different student characteristics. They were identified using snowball sampling techniques and approached based on their interest in evidence-based policy.
- Many teachers make use of research evidence for changing their practice, and several discussed research evidence from the Education Endowment Foundation. Teachers who frequently implemented research evidence emphasised a combination of factors of success, including understanding the underlying rationale or causal mechanism for a particular approach and fidelity to the evidence-based approach. In line with previous research, understanding how contextual factors might interact with an intervention was essential.
- Some teachers were sceptical of the capacity of formal research evidence to inform their practice, one suggesting a more 'holistic' scholarly approach to evidence was required rather

than a summing up of research studies. The capacity of individual teachers to alter individual practice varied significantly. Some had significant autonomy in their own classroom while others deferred to senior management. Time constraints were a key barrier to teachers engaging more carefully with research evidence.

1.3 Policy Pointers

- School leaders might consider setting aside more scheduled time for teaching professionals to engage with research evidence, perhaps as part of a reformed programme of Continuing Professional Development. Policymakers could facilitate this by reducing administrative burdens that can curtail time formally given over to CPD.
- Government bodies, especially Ofsted, should generally avoid setting policy requirements that contradict the research evidence provided by the EEF. Setting policy centrally that is inconsistent with research evidence defeats the aim of encouraging educational professionals to examine their own practice in their local contexts and of making use of the evidence.
- Institutions that host research evidence, like the EEF, should provide resources specifically aimed at supporting the local implementation of evidence-based policies. This might be achieved through having a dedicated implementation team, or by encouraging schools to share implementation knowledge as peers. There should be more opportunities for interaction and feedback on websites hosting the research evidence.

2. Introduction

2.1 Where we stand

The dissemination of evidence-based policy (or evidence-informed practice) across public services has a unique potential to improve public service provision (Sharples 2013). In 2002, the Centre for Evidence Based Policy and Practice summarised what was needed for successful evidence-based policy into four key requirements:

1. Agreement as to what counts as evidence in what circumstances

2. A strategic approach to the creation of evidence in priority areas, with concomitant systematic efforts to accumulate evidence in the form of robust bodies of knowledge

3. Effective dissemination of evidence to where it is most needed and the development of effective means of providing wide access to knowledge

4. Initiatives to ensure the integration of evidence into policy and encourage the utilisation of evidence in practice (Nutley, Davies, and Walter 2002, 2).

At the time, it was noted that research in school education was beset with a number of challenges, including research 'paradigm wars', 'eclectic methods competing rather than complementing' each other, and while there were some analyses of large datasets, there was 'relatively little true experimentation'. In addition, the education research community was described as 'fragmented' with 'no accessible database of research evidence' and 'few systematic reviews' (Boaz et al. 2002, 7).

Since then, significant progress has been made. Initially supported by more general evidence clearing houses, including the Campbell Collaboration, the UK education sector now has a dedicated What Works centre in the form of the Education Endowment Foundation (EEF), which also funds a large number of field trials across the school system. Hence, the EEF has:

(1) reduced the fragmentation of research with an easily accessible toolkit of evidence;
(2) produced more systematic reviews of existing evidence; and
(3) provided impetus and support for high-quality research experiments in schools.

The issue of competition and conflict over methods has not been extinguished (Goldacre 2013b; Biesta 2007; Phillips 2005) but this is not necessarily a barrier to progress. We can reasonably expect such a complex and contested policy area as education to provoke debate over the correct methods of establishing – and of using – evidence. That debate and contestability, as well as the inherent fallibility and defeasibility of any evidence-base, can be acknowledged, even as policymakers and practitioners act on the research that is currently available. Final agreement as to what counts as good evidence may not be obtainable and may not even be desirable.

2.2 The challenge

This report focuses on an aspect of the fourth point of the scheme of requirements for successfully implementing evidence-based policy. We want to work out how education professionals can integrate What Works evidence with knowledge of local circumstances and other kinds of available scientific knowledge to predict if the policy will work in a particular setting and to recognise what other kinds of information they may need to do this reliably and where they might find this information.

First, we review some of the existing literature (especially systematic reviews) on the challenges of implementing evidence-based policies, highlighting where possible the knowledge gap between 'what works' and 'what works here'. Second, we review the available information from the EEF on implementing policies in local contexts and compare it with other similar information produced by *what works* evidence clearing-houses. Third, we introduce new data based on interviews discussing how

teachers and other practitioners have tried to implement evidence-led practice in school settings and the challenges they have faced. Finally, we offer some proposals that may support the better implementation of evidence-based policy.

3. Existing Literature and resources

Our literature review comprises four sections. The first outlines some perspectives in the theoretical debate surrounding evidence-based policy, especially in medicine and the health sciences where the debate is most advanced. The second section examines existing empirical evidence, especially systematic reviews, of implementing evidence-based policy, with emphasis on implementation in education settings where available. The third section looks at literature focussed on practical advice for implementing evidence-based approaches. Finally, we examine the resources provided by the EEF.

3.1 The theoretical debate

Evidence-based medicine (EBM) and, in the United Kingdom, the National Institute for Health and Care Excellence are often taken as a paradigmatic approach to interventions in other domains. So it is worth noting that even in medicine and the health sciences, EBM is subject to frequent criticism in the academic literature. Indeed no sooner had evidence-based medicine been launched than it was accused of abusing human participants (Harford 2014) and of inherent bias against some forms of interventions and in favour of others (Sheridan 2013).

As an illustration, David Healy, writing about evidence-based medicine in mental health claims 'RCTs and the embodiment of evidence derived from them in guidelines have become a solution for complexity and a substitute for wisdom and in some cases for common sense (Healy 2001).' In his area, Healy is concerned in particular with research processes being captured and gamed by special interests, particularly by pharmaceutical companies. Of course, proponents of evidence-based medicine (EBM) and policy (EBP) hold many of these same concerns. Indeed, some argue that it is precisely a rigorous requirement for evidence that can challenge vested interests (Goldacre 2009; Goldacre 2013a).

Apart from concerns about the integrity of individual items of research, there is concern that research programmes and agendas can be led by demands from policymakers and the interests of powerful groups (Smith, Ebrahim, and Frankel 2001; Greenhalgh et al. 2014) or by other arbitrary factors (Young et al. 2002; Parsons 2002). In the area of drug addiction, for example, there is a widespread belief amongst proponents of psychotherapeutic interventions that the evidence base is too focussed on pharmaceutical interventions because pharmaceuticals lend themselves to testing with RCTs far

more readily than psychotherapeutic interventions and because researchers with the skills to conduct RCTs emerge from clinical medicine (Cowen 2012).

More radical voices have used critical theory to challenge what they see as a new form of power, rather than a genuinely transcendent scientific practice. One group of health scientists argue: 'the evidence-based movement… is outrageously exclusionary and dangerously normative with regards to scientific knowledge (Holmes et al. 2006, 183)'. They suggest that the 'norms [of evidence-based policy] institute a hidden political agenda through the very language and technologies deployed in the name of "truth" (Holmes et al. 2006, 183).'

Some of these concerns and criticisms might be described as excessive. Nevertheless, they indicate that even in clinical medicine and the health sciences, there are dissenting voices. All academic disciplines produce some 'push-back' when faced with a drive towards evidence-based policy. Many of these same concerns are likely to be reflected by professionals and practitioners in the field.

A strain of these criticisms in the sceptical academic literature could be described as 'interpretivist' in contrast to 'positivist' or 'realist' in approach. Interpretivists tend to reject the possibility of subject- and context-independent knowledge in the social (and sometimes natural) sciences in favour of 'sense-making' through socially constructed beliefs and practices. This research paradigm tends to be sceptical of the truth, efficacy and often the desirability of evidence-based policymaking, primarily because it 'privileges' certain beliefs above others in ways that are undemocratic and that cannot, they argue, deliver the kind of generally reliable scientific knowledge promised.

As a result, some of these discussions tend to problematize the implementation of evidence-based policy in local contexts without offering many prescriptions for increasing the chances of successful implementation. However, there remain important lessons to be drawn from this interpretivist literature precisely because it is focussed on pitfalls. In being concerned about how practitioners in a policy field 'make sense' of academic research evidence, they are examining a part of the evidence eco-system that more 'positivist' researchers can overlook because they often lack the tools to analyse it.

Indeed these criticisms have prompted useful defences of EBM and EBP that also help to clarify the challenges that EBP faces. Cookson (2005) argues that good use of evidence-based policy involves making use of non-scientific information:

> [A] broad range of theoretical and empirical evidence about human behaviour may be relevant to predicting policy outcomes, including stakeholder opinions and other sources of intelligence that might not qualify as scientific research… [O]bservational data, modelling and other less "scientific" forms of evidence are relevant as well as interventional trials…(Cookson 2005, 119–120).

In addition, although evidence can inform action, it does not set a baseline for action or inaction (it may be necessary to intervene even if evidence for the success of any of the possible ways to do so is absent) and values and principles also inevitably guide practice. The evidence cannot be followed blindly:

> [P]rior beliefs about the effectiveness of interventions are crucial in policy making. When evidence is lacking, it is quite proper for policy-makers to forge ahead on the basis of their beliefs… [E]vidence about outcomes is not everything… [for example] it may be wrong to allow age discrimination in health care even if this would slightly reduce inequalities in lifetime health outcomes (Cookson 2005, 120).

These caveats help to map out some of the additional kinds of knowledge that practitioners may need to implement policy successfully.

Cartwright and Munro (2010) introduce some important concerns about how to interpret randomised controlled trials, often considered to be paradigmatic of *What Works* evidence. They argue that randomised controlled trials by themselves are 'insufficient to meet the needs of policy or practice decision makers (Cartwright and Munro 2010, 265)'.

They point out the primary purpose of an RCT is to make a judgement on effectiveness where the contribution of other circumstantial factors are unknown (and hence can only be controlled through random selection in a research design): 'it is widely acknowledged that we generally don't know all the important causes for a factor, let alone knowing the distribution of subpopulations homogeneous with respect to these in the study and the target populations.' But since the same policy or programme will have different effects in different subpopulations – sometimes ranging from highly positive to highly negative, for new settings what matters is which subpopulations are present there and in what proportions. So extrapolating from positive outcomes in an RCT (or even a set of RCTs) to the claim that some intervention will work in a new setting or that it works in general assumes precisely a level of knowledge that RCTs are designed to cope without. If that level of knowledge were already present, an

RCT would hardly be necessary; since an RCT is necessary, that knowledge of generalisability to different circumstances is likely to be absent.

Cartwright and Munro trace three stages of knowledge that allows an RCT to be relevant to an intervention in a specific setting:

> *1. It-works-somewhere claims: T causes O somewhere under some conditions (e.g. in study population X, administered by method M).*
> *2. Capacity claims: T has a (relatively) stable capacity to promote O.*
> *3. It-will-work-for-us claims: T would cause O in population q administered as it would be administered given policy P (i.e. effectiveness claims)* (Cartwright and Munro 2010, 262)

Their key argument is that a move from merely claiming 1 to stages 2 or 3 requires an understanding of the causal mechanisms underlying an intervention. They acknowledge that 'this kind of complicated causal reasoning is hard, even if we are prepared to be rough in our approximations and figure out ways to tolerate uncertainties (Cartwright and Munro 2010, 263)'. They suggest though that sometimes large effects can be relied on to overwhelm other, less observable, features of a model and thus can allow practitioners to act even without much knowledge of other factors.

It is important to note that existing RCTs in medicine rarely take a simple 'proof of the pudding' approach to an intervention either. There are generally many more measures taken on the patient than merely measuring whether the desired (or undesired) outcome occurs. Many intermediate results and metrics are taken, all to aid understanding of the causal mechanisms which will make a treatment successful or unsuccessful. Understanding *why* something works in all scientific fields is a substantial aid for understanding *when* and *where* it will work.

In the area of complex public health interventions, Bonell et al. (2012) propose a solution, alongside several other approaches, in the form of 'realist randomised controlled trials'. These trials would attempt to identify the complex causal mechanisms behind an intervention through the use of 'factorial trials', 'multi-arm studies with various combinations of intervention components in each arm (Bonell et al. 2012, 2303)'. This extension of multi-site and cluster trials certainly contributes to the solution. The weakness of this approach, besides the spiralling costs, need for ever larger sample sizes and the challenges of implementing and co-ordinating complex experiments, is that some factors will remain always untestable, unobservable and out of reach of study designs.

3.2 Empirical literature

Despite the importance of successful implementation for realising almost any benefits of evidence-based policy in practice, definitive research on successful implementation itself is scarce (Blase and Fixsen 2013a). There are some toolkits available in the US in the area of education and more so in other fields such as health. There are also a handful of systematic reviews, for example Fixsen et al. (2005), that provide a synthesis of implementation (we discuss themes from this area in the section on practical literature below). There are also a number of studies that cover difficulties of implementation that have relevance to this research question. One recent systematic review of both barriers and facilitators of evidence-based policy focusses on health policy settings but is not limited to that sector. The authors identify 145 studies using a variety of study designs, including ethnography and surveys, 'but the majority were in part or in whole based on semi-structured interviews (Oliver et al. 2014)'.

'Evidence' itself was not always well-defined in the studies, although 33 referred to systematic reviews, making at least some of the results applicable to the information disseminated by What Works centres. The review identifies a number of common themes, many of them illustrating the radically situated factors that can make or break implementation. Successful implementation was often due to 'serendipitous', 'informal' and 'unplanned' factors, including 'mutual trust' and good personal relationships between researchers and policymakers (Oliver et al. 2014, 4). Discussed in more than a third of the studies was the significance of 'informal evidence', including 'local data' and 'tacit knowledge'. The authors note 'Identifying these sources and types of information are a crucial step in describing and ultimately influencing the policy process (Oliver et al. 2014, 8).'

Another systematic review tries to identify political factors that helped or hindered implementation of evidence-based policy in public health (Liverani, Hawkins, and Parkhurst 2013). The review notes an interesting difference between centralised and decentralised government in a study contrasting the UK's NHS, prior to reforms in the late 1990s, with the US federal system. Neither situation suggested that the smooth implementation of purely evidence-based policy was a given:

> *[C]entralised political systems are likely to be less open to the uptake of research findings than de-centralised systems. The concentration of power was found to prevent pluralistic debate... By contrast, it was argued that in countries in which policy is made through ad-hoc, issue specific coalitions, such as the United States, and in federal systems in which policy is made at the provincial level, "there is more need for research as legitimation or ammunition" to justify policy decisions and defend them against the criticisms of opponents (Liverani, Hawkins, and Parkhurst 2013, 4).*

This conjecture garnered from the studies reviewed could be suggestive for the UK now, given that the school system in England and Wales is currently undergoing, along some dimensions, a process of decentralisation:

> … schools are granted much greater freedom to pursue their own approaches to teach[ing] … [but] with this freedom has come increased responsibility to make informed choices, as teachers and commissioners are faced with a myriad of different strategies and interventions to choose from, each with varying levels of effectiveness (Sharples 2013).

The review notes that some of the studies surveyed show that in the situations they investigated 'personal connections between knowledge brokers and decision makers can be an important factor increasing the uptake and use of evidence'; although in more developed countries (including the UK), 'greater attention was paid to the reputation and professional legitimacy of institutions charged with the production or use of evidence (Liverani, Hawkins, and Parkhurst 2013, 5)'.

It also warned:

> [P]olitical pressures may encourage a selective use of evidence as a rhetorical device to support predetermined policy choices or ideological positions, or may delay decision-making on contentious issues while less contentious topics with clearer, uncontested evidence bases are followed' and that 'even institutional bodies that operate according to the most rigorous procedures may be vulnerable to distortions and/or the influence of interest groups (Liverani, Hawkins, and Parkhurst 2013, 6).

Much of the published research on implementing evidence-based policy in school settings is from the US. Besides the institutional differences already mentioned, it is important to clarify what is meant by evidence-based policy in this literature. In short, it is quite broad. Some federal funding schemes formally require the use of social scientific research findings (Honig and Coburn 2007, 580), but 'evidence-based' reform also extends to use of student performance data and local programme evaluations (Honig and Coburn 2007, 581).

Data-driven approaches to school evaluation is, of course, already a mainstay of the British educational system (most prominently through published school league tables), but they are not considered as part of the evidence-base that is the focus of *What Works* (although the objectives of What Works centres are indeed driven by quantitatively measurable outcomes). As a result, the US research is discussing

'evidence' that includes, but is not limited to, academic research evidence. Honig and Coburn warn: 'Arguably using student performance data… raises… different… challenges than choosing… school reform designs based on social science research (Honig and Coburn 2007, 584).'

Honig and Coburn (2007) evaluate over 50 studies that considered the role evidence played in decision-making in School District Central Offices (equivalent to Local Education Authorities). They identify a wide range of factors impacting on the success of implementing evidence-based policy in the sites covered by the studies surveyed. Characteristics of the evidence itself were often important, including: 'availability, accessibility, ambiguity' and 'credibility' (Honig and Coburn 2007, 594). They concur with other studies that: 'Evidence use… frequently unfolds in social interaction, and, fundamentally involves interpretation'. The shared meaning of evidence, they hypothesise from their review, is 'profoundly shaped by… pre-existing beliefs and practices and day-to-day limits on… attention (Honig and Coburn 2007, 585).'

This review raises two key issues that are of particular significance to this report. The first is the role of local knowledge that practitioners in the study sites used to make sense of research evidence:

> …[D]istrict central office administrators use a variety of forms of evidence… beyond those formal sources… include[ing] what some researchers call practitioner knowledge or local knowledge. Critics might argue that these other forms of evidence are inappropriate or less valuable than social science research evidence and that reliance on these other forms is precisely the pattern that federal policy makers should aim to break. However, the studies we reviewed here suggest that these other forms of evidence may be essential to growing and sustaining school improvement efforts. Furthermore, practitioner knowledge may help district central office administrators use the more formal types of evidence that federal policies favor by giving meaning to information and suggesting viable courses of action (Honig and Coburn 2007, 601).

The second is the importance of tailoring research outputs to the needs of local policymakers:

> [P]olicy researchers may bear some responsibility for creating conditions for effective evidence use in district central offices. Policy researchers might consider how to better align their work with contemporary challenges in school district central offices including crafting research questions and presentations of findings that speak directly to district central office audiences (Honig and Coburn 2007, 604).

The Department for Education in the UK commissioned an academic review of the implementation of evidence-based policy in children's services (Wiggins, Austerberry, and Ward 2012). This report draws significantly on Fixsen et al. (2005) and applies some of the principles to a UK policy setting. They identify 'the importance of careful planning and an expectation that it will take considerable time and resources to embed these programmes.'

> Innovative solutions have been found to overcome cultural differences, language barriers, and different system structures. There are examples of very successful replication of the original programme's positive outcomes in new settings. Despite this, some programme sites have found implementation and/or the replication of original success unachievable; and, in other sites programmes have been successfully implemented but found to be unsustainable when reliant on mainstream funding. (Wiggins, Austerberry, and Ward 2012, 26)

To sum up, the existing literature suggests three key things of significance for this report. The first is that it may often be impossible to implement evidence-based policy in the absence of other kinds of knowledge. Local practitioner knowledge, as well as locally collected and interpreted data, may often be essential for guiding policy decisions. Second, the implementation of policy cannot be turned into a routine set of instructions and will inevitably rely significantly on the autonomy and judgement of those actually implementing a policy. The third is that, despite this, there are ways of writing research evidence that make it more practical and useful for local policy practitioners and professionals.

The literature also highlights important political dimensions to implementing evidence-based policy. It is probably not possible to extract the 'political' element in its entirety from the process of implementing evidence-based policy. It is unrealistic to expect evidence, even of the highest quality, to cascade smoothly from academic sources, through policy clearinghouses, into the classroom. Dealing with this issue of politicisation is not the focus of this report. However, we might tentatively suggest that successfully integrating *what works* evidence with knowledge of local circumstances might also help to ease inevitable tensions that arise when trying to implement policies in local contexts. A more evidence informed approach to implementation might not only be more effective but also more likely to gain 'buy-in' from key actors and stakeholders.

3.3 Guidance and implementation literature for practitioners

Guidance specifically for practitioners is limited and that which is available is largely from the US. One such guide was produced by the US Department of Education, Institute of Education Sciences, the National Center for Education Evaluation and Regional Assistance (*Identifying and Implementing*

Educational Practices Supported by Rigorous Evidence: A User Friendly Guide 2003). This sets out the critical factors in establishing evidence of an intervention's effectiveness and important factors when implementing such interventions. It also provides information on where to find the evidence-based interventions.

More recent publications include *Exploring the implementation landscape* (Blase and Fixsen 2013a), which reports issues to consider and the resources required. The National Implementation Research Network has produced an active implementation hub with resources including a planning tool for implementation (Blase and Fixsen 2013b).

There are some other resources that provide implementation support, while not a toolkit as such.[1] These include a module to work through when considering choosing a programme from the Iris Center, a national centre dedicated to improving education outcomes for all children, especially those with disabilities from birth through to age twenty-one. The Department for Education in the US has produced excellent practice guides on various topics such as teaching maths; the guides include the evidence base and provide recommendations.[2]

There are also specific resources for evidence-based practices (EBP) for those with special educational needs. Again from the US, the National Secondary Transitions Technical Assistance Centre[3] is a centre providing national technical assistance and dissemination by the US Department of Education's Office of Special Education Programs. It provides technical assistance and disseminates information to state education agencies, local education authorities, schools and other stakeholders to implement and scale-up EBP for improved academic and functional achievement for students with disabilities. It provides factsheets that include research and also lesson plans for practitioners.

Implementation guides from other areas of social policy can be used, for example in health by Houser and Oman (2011) which is a very through guide. This looks in-depth at various aspects from making the case for evidence-based practice, to choosing the programme, to implementation, including various aspects of implementation, such as mentoring. Various worksheets are available in the appendix to assist the implementation process.

One controversial issue about implementation that needs to be broached is programme fidelity, that is faithful replication of the programme that has been proven to work in some, or maybe many, places.

1 See, for example: http://www.promisingpractices.net/
2 http://ies.ed.gov/ncee/wwc/pdf/practice_guides/early_math_pg_111313.pdf
3 www.nsttac.org

Fidelity matters because when programmes are modified to suit site-specific preferences or circumstances, in essence a different programme results, a programme that no longer benefits from the evidence available that the programme in its unchanged form has worked well elsewhere. Also, some studies on specific evidence-based programmes have noted that in the sites studied these programmes have produced better outcomes when implemented with fidelity than when modified in various ways. And there is also evidence that fidelity can sometimes be achieved when implementation supports are put into place (Fixsen et al., 2005).

Some commercial suppliers of reading and other policy packages insist that the package be implemented exactly as the supplier says. This is understandable. Suppliers cannot be responsible for failure if their instructions are not followed: 'The cake won't rise if you don't follow the recipe'. Moreover there is always the temptation in implementing a policy package to try to substitute cheaper, easier or more readily accessible components for some of those in the protocol that was tested. Clearly this can result in undermining the power of the package to produce the promised results.

There are cases where the post-mortem on a failed programme does indeed identify infidelity as the problem. For example, the California class-size reduction programme of 1996 called on evidence from a study of what was taken to be a similar programme in Tennessee. But the successes of the Tennessee programme were not at all duplicated in California. One of the chief things that went wrong was that there were not enough good teachers and not enough classrooms available when it was implemented in California, whereas in Tennessee they got both of those right. So a faithful execution of the policy as it was executed in Tennessee may well have worked, infidelity was the problem.[4]

Cartwright and Hardie (2012) argue that this illustrates the importance of having in place in the new setting all the helping factors (like ample supply of good teachers) that are needed to make the policy work there. Otherwise something needs to be done to ensure these factors will be in place, which can often involve changes, sometimes very significant changes, to the original protocol. The advice "be faithful" to the study protocol presupposes that the policy protocol builds in all the support factors that will be necessary for it to succeed, which cannot generally be assumed one way or another. Also, sometimes it is impossible to put in place some of the factors called for in the protocol but it may be possible to find suitable substitutes. For instance, to get a homework programme to work where pupils have no appropriate work space at home, it may be possible to introduce an after school homework club that can play the same role even though that was not part of the original protocol (Cartwright and Cowen 2014).

4 http://www.classize.org/techreport/CSR_Capstone_prepub.pdf

Fidelity of Implementation: Selecting and Implementing Evidence-Based Practices and Programs[5] is a module available on-line which goes through the different stages that should be considered for implementation where fidelity is likely to be effective. It discusses the importance of selecting evidence-based practices and programmes. It also examines actions that school personnel can take to increase the likelihood that the practice or programme is implemented as it was designed, though it does not provide serious help in deciding when fidelity is best and when adjustment to local circumstances might achieve better results.

Many guides, like the one above, urge programme fidelity. But equally, there are many voices on the other side. For instance, economist Angus Deaton, who has argued vigorously for the need to understand the mechanisms by which programmes work in order to decide if they will work in a new setting, considers a caricature example involving two schools, St Joseph's and St Mary's. He imagines that a perfect RCT has shown that a new training programme improves reading test scores in 6-year olds in the sites studied by, say, X. St Mary's is thinking of adopting the programme. St Joseph's has adopted it but adjusted the details in ways they thought would better fit to local circumstances and St Joseph's got an improvement Z. What should St Mary's do? Should it adopt the programme and if it does, should it adjust it?

Deaton argues that it is not obvious, nor clear that St Joseph's is not a better guide than the RCT, nor indeed than an anecdote about another school. After all, there is in the information available no reason to think St Mary's is at the mean reported in the RCT, and it may be a long way from it. The mean for the unmodified programme is useful, Deaton argues, and should be considered, but it is not decisive. St Joseph's may be closer, more "like" St Mary's, and may have got similar results in the past. This, he proposes, is better than predicting for St Mary' an average over unlike schools. Perhaps, he urges, the board of St Mary's could go to St Joseph's and see the new policy in action. This is in line with Deaton's general advice that for best predictions we should try to observe the mechanism at work.

Two key principles for successful implementation emerge from this part of the literature:

1. *Proximity to the details of original programme as much as possible can matter. Equally, departing in just the right ways can be just what is needed to produce successful outcomes.*

Successful outcomes may depend on replicating the details of the original programme, for example if qualified teachers were used in the tested practice, then the use of student teachers may well not

[5] See http://iris.peabody.vanderbilt.edu/module/fid/challenge/#content

produce the same results. But they may depend on changing the original programme in ways that fit it better to local circumstances. There is no manual for deciding which is appropriate in any new case.

2. *Understanding the impact of local socio-economic factors*

During the planning stage the implementation team should carry out an exercise, deliberating and gathering what information they can in order best to:

- predict whether the evidence-based programme can be replicated in the locality or institution
- consider any factors which may or may not translate to this locality and whether this will hinder the results
- decide whether to implement the policy
- decide whether it is best to stick faithfully to the original tested policy or to make local adjustments.

As for all social policies and programmes, education interventions act in a socio-economic environment where there are lots of factors at play, many of which cannot be controlled for. Such factors include pupil characteristics, family background, school and neighbourhood environment. Key factors include the number of pupils learning English as an additional language and the number of pupils currently eligible for Free School Meals (at time of writing, a key poverty indicator but one that is now being phased out).

3.4 Education Endowment Foundation resources

The Sutton Trust / The Education Endowment Foundation (EEF) publish a Teaching and Learning Toolkit, a flagship evidence document developed by academics at the University of Durham. It is free and available online and its information can be filtered in a variety of ways, and also read as an off-line report. It provides summaries of over 5,500 educational research studies for teachers and schools on how to use the resources to improve the attainment of disadvantaged pupils.

Our discussions with teachers suggest that those who have viewed the EEF website have seen this toolkit first and foremost and tend to be highly complementary about it, noting that it is both attractive and colourful. It provides important and useful summary statistics of cost effectiveness relative to facts about the sites engaged in the studies they survey, evidence quality and estimated effect-size in study populations measured in months of student progress. The toolkit allows individual interventions to be ranked along these dimensions. The content is particularly successful at ruling out as effective on

average across study populations a number of popular interventions, which have a demonstrable lack of efficacy on average in the settings in the studies.

There are a few ways in which the headline presentation slightly differs from aspects of the content of the toolkit, which we can expect to be areas for development in the future. For example, an approach that is well evidenced to work in a good number of settings and is well evidenced to be cost-effective in these areas with a small positive outcome may appear from the headline statistics to be very valuable. However when viewing the document itself, it turns out that sometimes the effect is difficult to sustain. Positive interventions with an average effect of, for example, two months improvement often have this kind of overall result. This might indicate that the meta-analyses of studies (that is the systematic search and evaluation of all studies meeting a pre-established criteria) that the EEF uses find it difficult to distinguish relatively small positive effects from no effect. In academic research it is often the finding of statistically significant effects that prove valuable for publication (including of meta-analyses).[6] However in the real world, it is effect size (or 'oomph') that is often more significant for making policy decisions (Ziliak and McCloskey 2009). On some occasions, it is even possible that results that technically fail a statistical significance test might, due to their effect size, still be promising interventions in the real world. This could be shown in the tool by illustrating the small positive effects a bit differently from demonstrably large effects (for example, using a different colour code). This would indicate where visible gains are more likely to be found in the classroom.

In addition, breadth of results across study populations can indicate that schools should be more cautious in predicting mean results for their own case. But the variance of effect sizes is not straightforward to estimate, by contrast with the widely reported average effects size. Nevertheless, while it may be inappropriate to discuss formal measures like standard errors in the toolkit itself, it might be worth having some sort of indicator of the sheer breadth of variation in results in the treatment and control groups separately, which is easy to estimate.

The toolkit provides some useful indications of how pupil characteristics can interact with the interventions. However, these characteristics tend to be limited to age range and the relative disadvantage of pupils. Discussion of other student characteristics, including gender, ethnicity and language, are limited. There is relatively little information about where the evidence is developed. A typical comment is whether the evidence is mostly drawn from the US, which of course can make a big difference. There is as yet little indication about the environment of the schools, such as whether they

[6] It is increasingly acknowledged in statistical science that standard significance tests, though often helpful and indicative, can represent a somewhat arbitrary test of the credibility of an effect, with researchers acknowledging that the appropriateness of these tests various between research areas (Johnson 2013; Gelman and Robert 2014).

are located in urban, suburban or rural areas. Some of these individual and environmental characteristics could be decisive factors in producing a measured effect size. Showing the effects working in a range of circumstances is crucial in justifying the claim that the intervention will work widely.

Teachers may approach the toolkit with particular challenges and specific kinds of pupils in mind. They will necessarily approach the evidence from a particular school context. So some ability to rank or filter the available evidence according to its relevance to specific pupils will be of great use as it becomes available. It might not be necessary to display this information in anything like the same prominence as the summary statistics, especially in the short run, since it is more difficult to establish these kinds of results with the same rigor that we can currently establish that a programme fairly definitely does work on average in the settings studied. Nevertheless, it could be helpful to provide some additional clear visual indicators of where the evidence suggests that particular factors, whether environmental or pupil specific, are important or relevant.

The detailed documentation that accompanies the toolkit places an admirable emphasis on the importance of teacher professional judgement when considering these interventions. It could be further acknowledged that measured pupil achievement is not always the most immediate aim of a particular intervention. For example, the use of teaching assistants is not given significant value using the toolkit's headline measures (it depends a lot on how teaching assistants are used). However, it is possible that the benefits of using teaching assistants accrue to existing teachers partly through a better, less stressful, classroom environment. This may not cause a direct measurable immediate improvement on student achievement. But it may aid other helpful factors within schools such as teacher retention, which can lead to better student outcomes in the longer run. These additional factors can be quite hard to measure, especially using experimental designs, but it should be acknowledged that student achievement is not the only legitimate aim of each individual intervention, even if it is a final goal of an overall policy strategy. The evidence could also acknowledge the possibility of these less measurable effects on schools of particular approaches and policies.

Much of the EEF's research provides useful hints and clues for those seeking to implement evidence-based approaches in their setting. However, although they provide significant practical guidance on conducting trials in school settings in order to develop research, there is comparatively little documentation focussed specifically on implementing existing evidence-based approaches and in particular on how to 'weigh up' research evidence on how well a programme has worked elsewhere with local knowledge, with hypotheses about what the underlying mechanisms may be, with whether

these possible mechanisms can be expected to work in the local setting and with conjectures about what constitutes a sufficiently 'similar' environment to expect similar results as have been found elsewhere. Thus we have identified a potential gap, also noticeable in the international literature on evidence-based policy, between research evidence and practical advice for implementing it.

4. Gathering the new data

4.1 Research design

Since we are concerned to identify where gaps in current advice may be without insisting on strong starting hypotheses on where these are, we adopted an 'elite interview' approach to data gathering, meaning we selected interviewees 'because of who they are or what position they occupy' where the purpose is to 'acquire information and context that only that person can provide about some event or process' (Hochschild 2009).

Finding suitable participants to interview presented a challenge. Initially we approached the EEF, and some researchers associated with the EEF, asking them to suggest users of their evidence that were well-known to them, whether teachers, schools or local authorities. Relatively few suggestions were available and those that were put forward did not respond positively to requests for interview. As is often the case when engaging with an elite in public policy, we understood that speaking to researchers was not a priority for our target participants, and that a pro-active and flexible approach to accessing interviews was required.

An experienced researcher suggested that social media would offer a useful way of reaching out directly to the education sector. This proved fruitful. We found a number of potential interviewers commenting on education policy on twitter and writing articles on personal and group blogs. We approached them via email and several were willing to speak about the role of EBP in their practice and in the education sector as they saw it. Utilising snowball sampling, we were able to contact other suitable candidates for interview. Engaging with social media also revealed that some teachers had used connections forged over the Internet to establish their own informal conferences and workshops outside of the formalised structures of 'school leadership' and continuing professional development. They were also separate from conferences involving government departments and NGOs in that they were organised primarily by practicing teachers. We attended one of these teacher-led events in order to gain some additional context about how teachers use evidence when interacting primarily with each other and to look for more potential interviewees. The resulting sample is unbalanced, directly and intentionally in favour of those heavily engaged in practices and debates about evidence-based policy in the classroom.

4.3 Participants

Our search produced a helpful range of participants. Our sample included new teachers, young teachers with a few years' experience, more experienced teachers with some management role, deputy heads, head teachers and a school governor. They were based in a range of schools, including community primary and secondary schools, academies and one free school. Amongst secondary school teachers, subject specialisms included science, history, English, design and technology, and ICT. It included teachers who had qualified via the traditional PGCE route and through Teach First training. Our sample involved 22 individuals related to 12 separate school sites.

5. Resllts

5.1 Use of evidence

Teachers explained that research evidence, and the EEF toolkit in particular, had been helpful for influencing management decisions and participating in policy debates within school governance. One said he liked to 'Use it as a hammer when things like setting or performance related pay [come up]. Heads of department in meetings ask "do you want to introduce setting in year 8 or 9?" I've read quite a few studies on setting, but it's good to point to the toolkit and say "looks it's not worth our time or attention, it's actively harmful."'

A senior leader described how EEF evidence was used to change the way teaching assistants were used in the classroom:

> We have engaged with the Hattie research and outcomes in the Sutton Trust Toolkit to improve the effectiveness of our individual needs department. One headline in the research suggested TAs are not cost effective and do not add value to student outcomes. [It's a] Hard research message to deliver to very hardworking people. How do you tell colleagues that research suggests TAs may even subtract value? Our leadership responsibility was to re-emphasise to teachers that all students are their principle responsibility and it is their role to deploy all other adults in the class. No TAs should be permanently attached to an individual student. Some parents expect their child who has a Statement to have a dedicated number of hours and a personalised TA. Now we have subject specialist TAs. We have TAs who will stay in science and come to know the science curriculum very well help all students rather than follow one child from place to place.

Another emphasises how the EEF toolkit, in particular, was useful as an initial look at the evidence but requires interpretation in order to be applied: 'It comes with tonnes of different caveats. Use it not as guidance, but as a way of asking questions. Whatever diagnostic tool you are given, you have to use it to anticipate problems and think about whether you can apply it across schools.'

A specialist school teacher showed that EEF evidence could be widely interpreted. In this case, feedback was cited as important, from the EEF toolkit, but the lesson was applied to feedback for teachers, rather than students (alongside an innovative approach to observing performance):

> We are looking at performance management, quality of teaching and learning. We look at research. The most important part of a teacher's role is arguably feedback, as the Sutton Trust explains. Top of the list. How do we ensure that bit of research, on the pupil premium, becomes a reality in school here? We are looking at quality of feedback to young people. But then teachers, how do they know how they come across? Two of my staff have been active in an action research project, looking at CCTV or video in school. They've researched this, contacted some private companies and selected one. It means staff can look at their performance (in the class). They can show their performance to an expert in, autism, for example, and ask for some positive feedback (two ticks and a wish) to get people to think and reflect on how they can improve their practice.

When well-received by an individual teacher, as opposed to use in wider governance, deployment of evidence in the classroom does not necessarily follow. A young teacher explained: '[There is] very little influence I can have in terms of what interventions will be brought in. That is a decision for the senior leadership team. [It is] Only my first year. I should not be in a position of going "have you read this study, we should be doing that"; you get a bad rep. But I still see VAK (visual, auditory, kinaesthetic learning styles] on other people's lessons plans!'

However, a more experienced teacher indicated that they were afforded a large measure of professional autonomy, and this allowed them to use research evidence to adapt their practice more or less as they saw fit (although they had relatively little influence on their colleagues): 'I am fortunate to be in a school that gives me liberty to teach how I think best. It works well in this school, I don't know if that works elsewhere… I left one school because it was too prescriptive.'

5.2 Challenges to implementation

The additional demands of time and effort were identified as key barriers to changing classroom practice. This is so for senior management as well as teaching staff:

> … but my major problem is time – time to read the research evidence, attend conferences where there are examples of practice.

> Teachers do not have time, [there is] incredible pressure in terms of accountability. Often things become corrupted even when there is a nice idea and the best intentions. Very often because it's not implemented properly, or slowly enough, or embedded enough, it just ends up ticking a few more boxes.

> There is a huge gap still to be bridged – between academic research and classroom pedagogy; there is not a solution yet. All these things I'm invited to – I go with the mindset that I am a teacher, 90% class time with prep-time. When do I have the time to apply it?

> [It is] only because I went part-time, that I got the chance to develop some of these things, read things about teaching. [This is] not afforded to a lot of teachers. More would take it if they had the time.

> Don't do much on research. By the time you get to the frontline, its been watered down quite a bit. I don't sit about waiting for the next paper to come out.

Simple resistance amongst teachers was a related challenge. Pointing out weaknesses in someone's teaching abilities, even if well-founded, can impact on a teacher's self-esteem:

> An older colleague has an approach to teaching that is not just old-fashioned and out-dated, but sometimes detrimental to students. He was convinced the approach was working. The school supported him to change, but this challenges pre-conceptions about one's belief in [one's] teaching ability.

Another teacher, cited problems of 'Workload, access to information, supporting implementation, in some cases a lack of will, not the case in my school but in the teaching population as a whole.'

Compounding this problem is the lack of institutional incentives to engage with research evidence and use it to become a better teacher in contrast to the rewards associated with climbing the 'management ladder':

> [You could be the] best [subject] teacher in the world, kids loved your lessons, best pedagogical approach imaginable, [but] it wouldn't be recognised by the school compared to a teacher that winged every lessons, got by on charm. [It is] perverse that there is no system by which good classroom practice is rewarded.

It was also suggested that some aspects of teaching might simply not be amenable to evidence-based reform: 'I would stress, to me it's very clear that teaching is a bit of science and a bit of an art, almost sorcery!'

Another key factor, related to time, though only mentioned explicitly by a few of those interviewed, was funding. Time does equate to the requirement of resources and this is usually viewed in terms of existing staff, however, money for resources and potentially additional staff was also mentioned. One interviewee questioned where additional resources for trials or changes would come from. This applies to all institutions as value for money is paramount and any new initiative would come with risk.

One teacher suggested that it was not the implementation of interventions themselves that was the biggest challenge to implementation but identifying the problem in need of intervention:

> An intervention might actually be quite simple, just a little shift [in practice], but its about identifying the problem that is the hard part.

5.3 Context and supporting factors

Teachers were generally aware of the wide variety of factors that will affect the outcome of an intervention or approach. One emphasised the importance of parent reaction to successful implementation:

> It's ok learning about what works in different places, we then have to think how it would work in our own setting.

> [There is] pressure from parents in certain schools. You have to be more flexible when you have a wide spectrum of abilities. You have to follow local and school ethos. E.g. does the school use setting and streaming. Some parents prefer one approach to another. If you ask students to do independent work at home, do they have the time, space and ability to do that at home? There will be a lot of things helping to decide whether to proceed with something or not. [With] Middle class parents, you can try just about anything, they don't kick up a fuss. Here we have Asian parents and Muslim parents who would question the way we are changing things. They keep an eye on things.
>
> The ethos of a school, how results-driven, very large schools [would] have to be run differently from this school. Teaching style, subject – things that work in English won't worse elsewhere. Policies already in place at the school. We use a whole set of recurring processes that students respond to, so if you put in a policy that doesn't link with those existing processes, they contradict and students quickly notice that. That won't be as successful… An example might be how a behaviour management system and praise and reward system which did not co-exist. If a praise policy didn't link into that, it didn't fit together nicely. Instead of being sanctioned for bad behaviour, you get rewarded for good behaviour, that could confuse well behaved students who don't see it working for them.

Another said that something as simple at the time during which the lesson is scheduled has a significant impact on the measured effectiveness of an intervention: 'Given enough time, teachers could design research instruments to establish correlation or even causation between what they do and the outcome. But there are loads of confounding factors. E.g class on Friday works less well than class on Monday. It's fraudulent to turn round and say "as a result of technique x students are doing much better". '

A specialist school teacher described how dealing with some background factors, completely unnoticed by those without specialist knowledge or experience, could be crucial for allowing some SEN children to learn: '[A]n uncomfortable child cannot learn. You could have an autistic savant in this room but who wouldn't learn because this fan would be driving them mental.'

One teacher gave an example of where their practice and experience in their school deviated (in their eyes) from, for example, EEF research conclusions. The reasons for the deviation were the sheer number of children on Free-School Meals at the school. This meant both that one-on-one tuition (ranked as comparatively expensive given its impact) was both necessary and affordable in their context compared with alternative approaches:

> The EEF toolkit suggested one-to-one support is expensive, moderate impact. But we have found it's high cost, high impact. [We can afford it] because we have a huge amount from the pupil premium. We narrow the gap between FSM and non-FSM. We've basically bridged the gap, 5% either way in terms of stats. The EEF evidence does not equate to what we do in our school. [When it comes to] homework: the quality and the differentiated homework and students self-selecting homework can have a significant impact on progress. The problem here is that children [in this school] do not have safe home environments... So we are very cautious; yes it's great there is a model – but you cannot use that to tarnish everyone with the same brush.

One teacher pointed out that truly generalisable results tended to look vague, with the result that it was not clear, in all cases, how they deviated from current practice. There is always space for sceptics to question whether something specific can be applied in a local context:

> The argument is how do you even use evidence in schools? How do you make evidence generalisable, because it's so messy. How do you take an intervention done in an entirely different context and apply it in your school. No one argues against evidence in principle but people contest evidence when it goes against their own experience and will contest it on context grounds. By the time you try to isolate those variables, you end up with very broad principles like feedback. Who is arguing we shouldn't give kids feedback?

Some teachers echoed this point, saying that approaches with the strongest results in the toolkit were already quite widely accepted by teachers: 'We very much talk about metacognitive, we've talked about collaborative learning. Homework is a thing we are building on at the moment.'

> There is nothing terribly new in the toolkit, its all stuff that is known. Homework has been argued about a lot in the past. It is well regarded in the toolkit.

Teachers suggested that contextual factors rendered prediction difficult when considering most interventions:

> My personal view is its very hard to predict 100% what the outcome is going to be. You can quite confidently predict the impact of some things, but there is always an aspect of uncertainty. Can you increase the likelihood? You have to be adaptive. If you are using ideas from universities (I get stuff from twitter, huge resource base but that's not evidence-based really).

> How can I make it more reliable? I don't really know. It's one of the treasures of the education system that you are working with a group of people who you can't predict.

Another teacher drew an explicit comparison between knowledge in the natural sciences and the social sciences to suggest some limits to evidence-based research in education:

> One of the problems is that education research is ultimately social science, not like physics where we can definitely say we found something, or medicine where we've got a new compound we think does something and we do a double blind trial. I don't think education results work in quite the same way.

One teacher suggested a solution to some of these challenges. This involved both trying to match the approach to that actually described in research evidence and measuring intermediate outcomes to make sure, at least, that the approach was producing the initial effect required for the approach to work according to the evidence: 'I try to prove to my satisfaction fidelity to the research. E.g. for learning objectives – I show they have a better grasp of learning objectives than before. I did a little experiment, to see which approach produced better recall of learning objecting. If they recall LO, I hope the research indicates it will improve outcomes.'

Another approach to help overcome some of these challenges is carefully considering how large the sample size and how similar the context of the research are to the proposed intervention site:

> I guess a larger number of people involved in the study you are running is evidence it is more realistic. If I was implementing a strategy here, and it was tested from a similar style school with a similar intake, and similar curriculum, that would be more relevant to us.

One teacher indicated that fidelity was not always something to strive to maintain so long as there was understanding that one was departing from a model used in research:

> Research done beforehand would lend some weight to it. E.g. the Hattie research, meta-analysis that he carried out. [You give that] somewhat more credence than something mentioned in a [CPD] course. The question then is does someone really understand what they are doing. Or if it is x [intervention], they have some idea of x, but by the time you have applied it in context, its y. Its important that you understand what x was and why you are doing y now.

Another teacher argued that the huge range of contextual and supporting factors rendered experimental research evidence itself of limited practicality:

> I love how trendy [randomised controlled trials are] but what does it tell you? You can't negate the impact of one teacher's charisma, or one teacher's bad day because they are getting a divorce. And how are you measuring progress? I don't think randomising one method will ever give you generalisable results. I get to the stage where I start to think that kids aren't guinea pigs. There are things you can test in the human body and fix it. In education, there isn't one thing in different settings that you can reliably fix to get the same outcomes. Its not consistent in the way the human body can be. In schools, if you do different methods - is it the method? - or is it the teaching assistant? - is it about class, race or gender in the school? Not sure you can close down the variables enough for it to be that [useful].

5.4 Other sources of evidence

Interviewees that made use of the EEF toolkit tended to use it alongside other sources of evidence. One emphasised the importance of going to the source academic material, suggesting in particular the importance of understanding the mechanism through which an intervention is supposed to work.

> [The EEF provides a] very helpful introduction but not enough information to design a feedback policy. Really I need to look at the studies that have formed the meta-analysis and remember the flaws in meta-analysis. Take John Hattie, [who] aggregated primary and secondary homework policy into one effect size. [The toolkit] Doesn't provide all the answers, which is fine, it's not its role… You have to tailor it to your classroom. This is where reading the research is important. You have to understand that underlying rationale.

Another teacher argued that in order to interpret research evidence, it is necessary to have some grounding in theoretical frameworks:

> We are pushed to be teachers as researchers, when really we should be teachers as scholars. There is a theoretical aspect to teaching practice, not just classroom practice. As rounded practitioners, we should be thinking about all aspects of our work. A lot of people haven't read a lot of education theory, and we need to have both research and theory to read and understand. We need that in order to navigate around the various trials and say 'well this one is no good because they did this and gives me a load of figures that mean nothing in my context

> or whatever'. Just because someone has done an education degree from 18-21 [does not mean they have] a lot of experience of decoding research.

Teachers sometimes favoured particular academic pedagogies and had developed affinities for particular ways of understanding classroom practice and favourite authors:

> I use Dylan Wiliam's book, *Assessment for Learning*, as a bible… what he gives us, modified by Daniel Kahneman's work, is essentially a theory of learning [which], unlike Vygotsky and Piaget, is about how teachers can train the mind in practice, the importance of retention and focus. This helps you design a lesson to aid retention.
>
> Hattie fits into a pattern of ideas that is confirmed by other people. For example, formative assessment fits with Paul Black and Dylan Wiliam.

One recently qualified teacher described using Maslow's Need Hierarchy and Bloom's taxonomy, both introduced through a university course, and was starting to integrate SOLO taxonomy (introduced at an informal teacher conference) to structure their thinking and classroom practice.

Others were more eclectic, openly drawing on a variety approaches for their own practice: 'For any Willingham or Hirsch, you have some counter-evidence. Even this [useful approach] is dangerous. [It's] nice to look at different aspects of things.'

One teacher mentioned that their school had recently subscribed to a number of academic journals on JStor and that looking over academic papers on blogs and within the school's inquiry group formed part of their teaching development.

Action research was also cited as a source of evidence. Action research, often conducted as part of a Master's degree in education, typically involves a researcher-practitioner introducing a new approach and evaluating it using feedback such as a survey. It tends not to include an experimental element or systematic data collection, although it may include some sort of before/after comparison: **'**I would not take seriously studies with one teacher, one classroom. You have to take those with a pinch of salt.'

Reflecting our social media search strategy, the Internet was cited as a source of evidence:

> We don't engage with [research evidence]. [But] This stuff is changing, [I'm a] big fan of using twitter – twitter is a wonderful driver for self-sought CPD – it's the world's biggest staffroom. One in four schools don't have a qualified [subject specific] teacher but on twitter you can find them.

Another explained, when discussing CPD: 'I write a blog and read other people's blogs, and I experiment a lot with my own teaching.'

Developing out of social media discussion, a number of national and regional teacher-led conferences were also discussed.

Some teachers relied on informal small-scale trials inside the school to see if a particular approach was working: 'You can come up with a thousand and one ways of improving an outcome but I always wonder what the evidence is. So when a member of my team comes with an idea, I always think trial it first.'

Another principal of a vocational sixth college described hosting international exchanges in the development and delivery of their vocational programmes.

For one specialist school, clinical research and collaboration with hospitals was an important input: 'We have two or three children with very rare disorders (1 in 2 million). Some of our severely epileptic children, if they get the right support medically, could be transformed in the future. We work with research hospitals and consultants.'

5.5 Dialogue and sympathetic engagement

A repeated theme was the value of teachers engaging and discussing evidence for themselves rather than passive instruction:

> Because I found out research for myself, there wasn't the sense of shame or humiliation that happens when someone tells you are wrong. That's why teachers need to be supported to engage with research evidence for themselves.

One teacher, involved in organising teacher-led conferences, explained that they put a lot of emphasis on dialogue rather than top-down instruction, with plenty of teacher-led programmes alongside: 'We want to see more collaboration between teachers. [It's] not about academics coming in and saying 'this works, do this'.

Another teacher, also involved in online discussion and informal teacher-led conferences suggested:

> I would like to have communities that look at research critically, not to trash it because a lot is really interesting, but also think about the things that make it not so relevant to your setting. There is a real gap between theory and practice. We argue about what we are doing in the classroom all day long but we should debate the theory too.

5.6 Data

All participants made some use of quantitative outcome data as part of the deliberation for further actions. A governor explained that management meetings included discussion of 'anonymised individual child data' often when identifying pupils who are falling behind. This was complemented with qualitative feedback from the class teacher and the specialist inclusion manager.

No one was opposed to using formal data. Each placed different emphasis on its value, often contrasting data to 'gut feelings' or intuitions:

> [Evaluation] can't just be a gut feeling. We need a level of accountability.

> You need to use your gut. You can't let data overrule your gut… Suddenly [if there is too much use of data] there is no role for humanist interactions in school. That's a bit severe. But the question is how many steps you would take before you had that. Data provides indicators and informs, but it doesn't judge. It shouldn't be making the judgement.

> Data can be a really interesting in-road to conversations with parents, for example. It can really improve a child's attainment. But raw statistical data, without triangulation, is not helpful.

> It is really good to know the children; I have had arguments that it's just about data. But actually knowing what they come with [from primary school] and where they are going is actually very helpful. We add a lot of value and that's because we know a lot about where people are. Nationally, there are some shocking stats with children earning less [from deprived

> background when they get to the workplace]. Each child will be known individually, whether they are making expected levels of progress, so and so might not be making progress and we have that conversation. We turn it to a human side.

Interviewees suggested that it was almost inevitable that higher attainment would turn up in higher scores: 'If they get an A*, they are probably better at [specific subject] than someone getting an A.'

However, they also claimed that it was sometimes possible to achieve higher measured outcomes without necessarily improving attainment:

> I am not anti-level and anti-data. They are necessary and useful. But there is gaming and processing of data... If that's how a school lives or dies, teaching to the test won't give a good idea of what level the children are at. I can get them up to a level by teaching "adverbs". They should be using adverbs but that should be a natural product of loving writing, developing writing. Data encourages ad hoc lessons to achieve levels… [It's] Horrible just labelling children "3a".

Another teacher argued:

> '[There are] so many different skills you are testing, and I'm meant to sum that up into a single data point. Truth is I don't know what they will get at GCSE. Obsession with data is harmful to schools and to what we focus upon. [It's] Important students know enough to pass the exams, but also that they are functional citizens able to read and write. Whittling things down to the data is harmful and creates perverse incentives, for example the focus on D/C borderline. The amount of gaming that goes on... We are also bringing up humans and if we show that we will cheat to get them a C, we aren't showing them how to be good citizens.'

Teachers also noted that data should not be limited to assessment in a given year or even a career within one school. Absenteeism was an important early warning sign of problems emerging, which teachers might spot on their own but that a formal registration system can be used to identify. Longer-term outcomes, such as attendance-levels at the subsequent school, could reveal challenges that exam results themselves did not identify:

> In the past, we have had a bit of problem with resilience, children leaving with excellent grades but dropping out of 6th form. So we thought about learning to learn. We [wanted] children to

> able to cope with setbacks. We have got these relentless targets for GCSE [but] we [now] have independent learning days for year 7. We introduced more project style homework.

One teacher noted a potential paradox in their own thinking on data:

> 'I contradict myself in a sense. When I study research, I am looking for clear methodologies and outcome measures. Ideally randomised controlled trials. In my own work, I don't think there is a particularly good way of measuring my holistic approach. Nothing that will use the data that I have in my head or saved in my mark book.'

When it came to their own practice, they were confident that measured achievement would eventually support their approach but they relied very little on quantitative measures and instead on subjective feedback of what seemed to be working. However, when it came to research evidence, they wanted to see measured outcomes.

It was suggested that just because a school is engaged in research, including RCTs, this did not mean that they were bound to follow the data. One teacher explained: 'There is an RCT at another school I have heard about. They really like what's come out; it doesn't quite tally with the national outcomes but it seems to be working there, but they have no data [in yet] whatsoever.' Hence, it seems possible for subjective experience of an intervention to dominate measured data, even when the formal research design is meant to clarify and test those subjective experiences

It was also argued, when evaluating interventions in school, the data never speaks entirely for itself: [It's] Hard to say if [pupils] would have made that progress without the intervention. Some still struggle to make progress even with interventions. [We] can't tell that just from looking at the numbers.

Some teachers and school management seemed to conceptualise a data-driven approach as itself an evidence-based intervention. In arguing they needed evidence to inform and develop their practice, they understood evidence to be knowledge gained through assessment of students. Each assessment revealed how far a student had progressed and therefore what further teacher input was required to help the student. These had tended not to have studied research methods at university and had no idea about how to set up a research based trial of new teaching methods. They had never heard of RCTs

Data analysis and benchmarking was an explicit programme used in one school. In practice, this seemed to correspond to a combination of performance management and aspiration interventions:

> We keep data accurate so our interventions are accurate. The advantages of a data approach is everyone's got something to aspire to, they want to see a target grade on their assessment. For some, especially lower-ability students, low target grade can be demoralising, so we insist it is a lower bound target. It can be a spur for many students. Target grades are given to us via [education charity] based on wide range of things. [Its] not something we create in-house.

> The progress agenda has changed how schools think – the focus is not on a single baseline, but on improvement. If you set targets from previous set of data, you may cap potential progress with some students. They are with us for 5 years, poor performing students on arrival could perform exceptionally well with us. Its always worth getting someone a B rather than a C and an E rather than an F. The starting point of the system is up for debate and people find it hard to be benchmarked.

A specialist school teacher explained that for their students, ordinary outcome measures were often inappropriate, but that they had found alternative way of observing student progress:

> I look at trends, and correlations, rather than solid empirical stuff because we are dealing with kids way off a normal distribution curve. I look at what works for mainstream schools. We had been struggling to evidence progress, but we now have a program that maps where the children are, with little performance indicators. E.g. can a child blink and recognise a face, or for one child, pick up a pen in one grasp. We can add on movement profiles. Some are well developed by other specialist schools. We use their descriptors and work on progress. We sample every 6 weeks or 12 weeks. And see about measurable gain and where they are at.

5.5 Ofsted

Ofsted emerged as a consistent theme. It was suggested that Ofsted can play an important role of auditing standards in schools. A school governor described one Ofsted report as a 'wake up call' in their school where 'Teachers didn't recognise [that] What was good in 1990 is no longer good'. Another role of Ofsted, from the governor's perspective, was as an evaluator of classroom practice for which governors are not equipped: 'I can't question [a teacher's] professional judgement, that's why we have Ofsted.'

One teacher saw Ofsted as part of a systematic drive towards accountability that was 'Good in general' so long as accountability was 'done in a fair way and moral way' with staff and schools given lots of chances to improve.

However, when it came to judging classroom practice, many teachers themselves often considered Ofsted more of a barrier than an enabler of evidence-based practiced. It was argued that Ofsted has endorsed non-evidence based practices (the paradigm being visual, auditory and kinaesthetic learning styles): 'Fear of Ofsted drives … a lot of dodgy research. [E.g.] Brain gym – there is a simple solution for everything and it's wrong. The magic bullet [that] will solve all your problems… I've had learners say I am kinaesthetic [referring to learning styles]. But it doesn't exist! We would all like to dance about, but we have to read.'

It should be noted that another teacher found learning styles useful and was not aware that this approach was challenged by research evidence.

There was a perception that classroom approaches that work well in practice (and in moderation) can be penalised by Ofsted inspectors, especially lessons that use a significant amount of teacher-led instruction:

> Chalk and talk all the time would lose your children. [But] Sometimes they are very happy with new learning. For example, I did square numbers today. It's in the maths scheme. I taught it to them putting it up there, there were heads down and they loved doing it because they knew it was difficult, because it looked different. Not a good lesson by Ofsted standards!
>
> In order to avoid censure from Ofsted, they apply things they think Ofsted like but don't necessarily understand… People look for things that will tick Ofsted boxes.

It was suggested that years ago, school inspectors were more in touch with research evidence: 'There was a time maybe 15 years ago when HMI issued best practice reports.'

At the same time, it was suggested that Ofsted was not necessarily directly responsible for some of these practices and that it was sometimes a school's eagerness to please them that led them to apply guidance in counter-productive ways. It was more that Ofsted had not done enough to challenge this reaction to inspections:

> Schools are a funny old bunch, we do unto ourselves sometimes – judging a whole school on individual lessons. We apply the observation model at the wrong level of abstraction – one

> teacher, one lesson – it's meant to be a whole school judgement. Ofsted don't bring out enough guidance to dispel [this myth].

One teacher suggested that Ofsted were willing to acknowledge the value of different approaches, and different outcome measures, so long as the case for them was made effectively during an inspection, although making that case itself requires skill and experience of the inspection regime:

> To some extent, it's about arguing with the inspector about what counts as progress. I've been inspected heavily since the late 90s. Some of my staff are Ofsted trained observers. We are familiar with the game that we need to be played. We need to make sure the data is true and staff aren't just trying to 'pass go'.

Another teacher indicated that they had personally shifted away from trying to fit their lessons into an Ofsted framework and that schools were increasingly judging performance in internal reviews differently from single graded lesson observation:

> [I] got very upset with graded lessons observations, something so stressful with so little formative value to staff… The lesson observation sheet is based on an Ofsted spec from 3 years ago; you have to observe progress in one lesson. We know learning doesn't work like that. In terms of attitudes, it was a shift for me, from [a] nonsense comment [in a review] to working out why Ofsted were trying to see you progress in 20 minutes. You can't observe learning, its invisible, so you can only observe proxies for learning. [I read an] article last week saying [local school] has given up lessons observations. This is just what most schools do, because they think this is what Ofsted need.

One teacher saw some potential for using Ofsted to encourage the uptake of evidence: 'The quickest way to effect change is to tweak what Ofsted are looking for. If you got Ofsted using evidence, people would take it seriously but that doesn't seem to be the case.'

5.6 CPD

Continuing Professional Development was described as in a state of flux. Interviewees consider it a potentially valuable way of engaging with research evidence. However, existing 'traditional' practice with occasional INSET days and outside consultants was considered inadequate. In its place, several interviewees discussed an emerging approach of shorter, more frequent, CPD sessions managed in house:

> Sending people on courses does not really work. It is about what you can do on-site to embed practice.
>
> CPD is just about handing data and assessment. I am shocked at the low level of expectations.
>
> [CPD] had a bad rap recently, [because] it's a 'done to' model, 5 bog standard days, everyone sits in a room and gets told what to do; never a revisit. I teach 193 days a year, 5 days training, no time for consolidation, then external costs. Social media is exposing poor practice and high costs… There are providers that are selling progress in 20 minutes, [which] sounds great, but totally against current practice. Selling the myths of education, that a child learns something in 20 minutes…. Schools need to think very creatively about CPD. [In a new school] we have removed one lesson of the day a week for training 29 lessons rather than 30, so two-hour session a week to embed and share action research and discuss problems. More useful model than 5 INSET days a year.
>
> We [now] get hour and a half [on a regular week day], where we send the kids home and we do professional development… Most teachers are very sceptical of CPD. They have had time wasted by charlatans and senior leaders who don't understand the techniques... I try to avoid bringing in external training – if anything, you should send someone OUT and bring the info back in, so you can tailor it to your setting. Not one-size fits all.

Another institution also had CPD sessions each week on a range of topics; the head also observed teaching and made recommendations to improve practices.

5.7 Encountering EEF research evidence

Descriptions of encounters with the EEF were mixed and varied. One new teacher, currently on the Teach First program had strong praise: 'I was really excited when I discovered the EEF. This is like a shortcut to really good quality evidence, like meta-analyses, the juicy stuff. Which means that your VAKs [learning styles] do get thrown out.'

Others explained that it was having some influence and was achieving some penetration within a sub-group of teachers: 'The Document [EEF toolkit] has done the rounds quite well in 'informed' educational circles, people [are] taking an interest'.

> The toolkit, that I am familiar with, is very useful. As a school, we have taken on a few things. Learning to learn (metacognition). We have certainly taken [that with] peer reviewing and how do you know what you know.

Others were more critical of their initial encounter:

> I've always had a problem with the toolkit – the way it was introduced. The first thing people said was "alright, lets get rid of TAs". Obviously that's not what the toolkit was saying but because they dropped it on schools like this, everything gets misinterpreted - no time to take it in with all its caveats. It was released on their blog – there are lots of tricky discussions about the role of TAs, why impact of TAs was so low because we weren't using them correctly. We need to ask how we are using TAs.

Others criticised the general lack of teacher involvement in dissemination:

> I've looked at EEF much more this year [than previous years]. But it tends to come through on twitter. But [it's] very hard to keep up with everything. Frustrating for think-tanks and policymakers to think they are suggesting a way forward without any classroom practitioners on the project.

> I always like it when institutions make things. That's lovely. But I wonder if they accept that people [can't] use it wholesale. [There is] not a particularly effective feedback opportunity. No Google + community. It's just a product. Those things presented as perfect are probably not. And we should think about where they come from.

One teacher interviewed because of their close engagement in research evidence for classroom practice had not yet encountered the EEF or its evidence, relying instead on their own reading of academic research and subject-specific debates about classroom practice.

6. Discussion

6.1 Fluidity, not fidelity

A key finding is that schools do not tend to apply evidence-based approaches 'out of the box'. Instead, new research evidence interacts with existing knowledge and practice in unpredictable ways. Evidence can be used as support for existing practices that get re-interpreted in terms presented by evidence. It

can also be used to change existing policy. For example, evidence that using teaching assistants can produce poor outcomes has led some schools to re-evaluate how teaching assistants are trained and where they are used in the classroom. In such cases, evidence is sometimes deliberately deployed to overcome institutional inertia.

Evidence has also been used creatively, new approaches inspired by research evidence but not actually supported by it. For example, evidence that feedback helps students might be cited in support of a feedback policy for teachers. From a classical evidence-based policy account, some of this activity could be problematic. These approaches to the evidence jump far beyond the available data. On the other hand, if it's accepted that an important role of research evidence is to develop hypotheses and provoke deliberation rather than direct practice, then these sort of approaches can be seen in a positive light.

6.2 Internal validity

Our initial research question was about the problem of what is often called external validity, or the challenges of generalising evidence of what works to different contexts. However, what is arguably the more basic question of internal validity came up in a number of interviews. This is the problem of whether a study or piece of evidence actually measures what it purports to measure. Here teachers, in particular, highlighted a potential problem. 'What works' studies tend to focus on quantifiable data, especially in the form of standardised tests. However, standardised tests do not necessarily measure underlying attainment. It's possible for an intervention, or change in approach, to improve a test outcome without improving the attainment that the outcome is meant to measure.

In addition, there are outcomes, such as student engagement in learning and student well-being, that are very difficult to observe and are rarely observed in formal quantitative research evidence.

6.3 Debate, dialogue and disagreement

Amongst interviewees, the role of evidence seemed inextricably linked to debates within the teaching profession, sometimes conducted over social media. A division between 'progressive' and 'traditional' education was mentioned and discussed in a number of cases. Interviewees rarely identified themselves as being in one ideological 'camp', suggesting instead that they drew on both approaches in their own practice. However, the debate itself seemed to be an important source of framing research and the underlying rationale of particular approaches. The debate also represented a source of motivation for further engagement with research evidence and a way of filtering research. When 'allies' (or indeed opponents) cited evidence in a debate, it might be sought out, analysed for corroborating or rebutting

claims. Some teachers were encouraged to engage, in particular, with recent research in cognitive psychology as a result.

Teachers seem to be interested in understanding the underlying mechanism through which an evidence-based intervention is supposed to work. In developing a mental picture of that mechanism, they tend to draw on a wider framework that fits some of their experience, intuitions and values, and this may include some underlying ideological pre-suppositions. It should be noted that ideology in this sense does not necessarily correlate to any particular political ideology, although perceived links are certainly present.

This disagreement over values and underlying frameworks means that the significance of even valid and reliable evidence is likely to be subject to contestation. There will not come a moment when all teaching professionals will interpret the evidence in exactly the same way. In fact, we might go even further and suggest that evidence-based approaches might not represent a consensus but instead a source of conflict within the education sector. By clarifying what is at stake in a particular approach, evidence-based approaches might bring the question of values into sharper relief, potentially a deeper source of disagreement than observed outcomes.

This conflict might in itself be nothing to fear, especially if it results in raising the quality of discourse and deliberation about education. We can point to the many antagonistic debates that take place within the medical profession over a variety of issues and suggest that, on balance, such a debate is probably productive even if passionate and occasionally aggressive. However, this might not be precisely what some proponents of evidence-based approaches have in mind as a desired result.

Interestingly, although the fast-track approach to Teach First is controversial and sometimes seen as de-skilling, our research suggests that some people entering teaching on the Teach First programme might be more likely to be engaged in pedagogical debates and research evidence. This may reflect their level of graduate education and, in some cases, subject specific knowledge and interests.

6.4 Institutions and accountability

If debate and dialogue is closely linked to teachers understanding of research evidence, then the role of institutions and accountability mechanisms is complex and often problematic. Generally, just because an evidence-based approach was mandated or supported by official policy did not mean that it would be effectively implemented. In fact, it could even mitigate against it being effectively implemented. Related to concerns with the validity of data-driven approaches to accountability, such mechanisms can

force teachers and senior leaders to focus on demonstrating effective implementation using bureaucratic criteria, potentially at the expense of implementation that serves the interests of students.

This chimes with recent existing research on education policy, which compares accountability and governance mechanisms in England with Scotland. Ellis (2007) and Ellis and Moss (2014) use the example of phonics in primary literacy education to suggest that the centralised nature of policy-making in England can lead to support for particular classroom practices becoming unnecessarily politicised. By contrast, the Scottish system of more local autonomy created an environment in which innovative research was possible and allowed for more reasoned interpretation of the results of that research. Ellis (2007) suggests: '…differences in the way literacy policy is determined in Scotland, and a greater knowledge of the wider funding and policy context, can explain the more measured response from Scottish policy makers' (Ellis 2007).

Ofsted consistently emerged within this theme. Although our small sample size limits our ability to generalise and differentiate, it seemed that the value of Ofsted oversight varied according to circumstances. Well-established schools with balanced intakes and a minority of children on FSM seemed to enjoy a productive relationship with Ofsted in which relative autonomy of senior staff and teachers facilitated the deliberation and implementation of effective evidence-based policies. By contrast, recently established schools, and those with challenging intakes, seemed to be more beholden to Ofsted. They were more narrowly focussed on meeting the needs of the inspection regime and this may restrict their ability to think more deeply about the implementation of evidence-based approaches. This chimes with concerns that Ofsted's national criteria for school judgements disadvantages schools in some neighbourhoods and with research suggesting that accountability mechanisms have often helped establish (repeatedly) where weaknesses lie without facilitating improvement (*The National Strategies: A Review of Impact* 2010).

It should be noted that some approaches to accountability, especially if they take the form of arbitrary demands for measured improvements in outcomes within a certain time-scale, actually subtract from an environment where research can be gathered. Evidence-based policy assumes that practice improves partly through a process of trial and error, simply because what works in a particular context cannot reasonably be known in advance. In other words, failure is an ever-present possibility that should be identified and amendable rather than something that is assumed can be avoided given the right practice.

The EEF, by encouraging and supporting experimental research in schools, could itself be contributing to a positive environment where trial and error learning is acknowledged and appreciated. But schools

may still have to tread carefully around existing accountability mechanisms when involved in trialling new approaches. It could be worth looking out for when the demands from such mechanisms might be contributing to schools dropping out of experimental trials.

7. Policy recommendations

A key theme was the time constraints of a full teaching timetable and the resulting challenge teachers face in trying to read and interpret evidence. Our data indicate that this problem is already being addressed in several schools by setting aside a period a week for CPD. One avenue for schools and the sector to consider would be increasing the number of teaching staff in order to leave more time for professional development. Up until now, increasing teacher numbers has been associated with policies of reducing class sizes. It is now argued that this policy is often quite expensive for a relatively modest measured impact. However, it is possible that using increased staff numbers instead to extend time for class preparation, and research and development, could produce durable improvements in classroom practice. This could be a legitimate area of additional expenditure to test.

Another theme was contradictory messages regarding research evidence emerging from different government institutions. To address this, it is not necessary for other government departments and institutions to march in lock-step with, for example, EEF evidence, not least because the evidence is tentative and still subject to many caveats. There is a wide range of justifiably good practice, especially given the difficulties of establishing the relevance of research evidence for particular contexts. A looser, perhaps more obtainable aim, would be to ensure there is some level of consistency between the evidence provided by the EEF and guidance provided by other government sources. For example, if a particular approach has been researched in some detail but lacks evidence of efficacy (individualised learning styles seems to be a potential example of this), it would be helpful if government sources could refrain from continuing to show support for them. Practice does not have to be proscribed as such in order for official recommendations to have some level of coherence and consistency with research evidence.

In terms of addressing longer-term issues, our interviewees highlighted the lack of pedagogical research training in PGCEs, as well as other channels into teaching. This relates to wider issues of qualification and expertise in the teaching profession. For the purposes of this report, we are concerned with making teaching professionals capable of interpreting and evaluating research evidence so that it can be applied appropriately in a local context. We propose giving trainee teachers the opportunity to become familiar

with research methods, including how to interpret statistics, and some understanding of causal mechanisms, in order to understand how to interpret and critically evaluate scientific research.

Funding, be it within that allocated to the development of resources such as the EEF or as a separate initiative, should be allocated specially for the implementation of policy. Otherwise the use of the resources will not be maximised and such resources will remain as banks of information rather than information used and improvements made as a result of the evidence. Resources which are solely information providers and lack implementation guidance cannot be considered good value for money. Implementation can be improved through funding for assistance with implementation, by, for example, a central team from organisations who develop the evidence base such as the EEF, or more local training in the use of evidence and availability of implementation guides and peer learning. Learning can be shared among those who are implementing similar practices as lessons learned. This can be carried out in the form of sharing practice on EEF-type sites or on web-forums. Interviewees suggested a number of times that sites hosting resources should allow for more interaction, encouraging teachers to provide and publish feedback.

References

Biesta, Gert. 2007. 'WHY "WHAT WORKS" WON'T WORK: EVIDENCE-BASED PRACTICE AND THE DEMOCRATIC DEFICIT IN EDUCATIONAL RESEARCH'. *Educational Theory* 57 (1): 1–22. doi:10.1111/j.1741-5446.2006.00241.x.

Blase, Karen A., and Dean L. Fixsen. 2013a. *ImpleMap: Exploring the Implementation Landscape*. University of North Carolina Chapel Hill: National Implementation Research Network. http://implementation.fpg.unc.edu/sites/implementation.fpg.unc.edu/files/resources/NIRN-ImpleMap.pdf.

———. 2013b. 'Stages of Implementation Analysis: Where Are We?'. National Implemention Research Network.

Boaz, Annette, Deborah Ashby, Ken Young, and UK ESRC. 2002. *Systematic Reviews: What Have They Got to Offer Evidence Based Policy and Practice?*. ESRC UK Centre for Evidence Based Policy and Practice London. mhttp://www.kcl.ac.uk/sspp/departments/politicaleconomy/research/cep/pubs/papers/assets/wp2.pdf.

Bonell, Chris, Adam Fletcher, Matthew Morton, Theo Lorenc, and Laurence Moore. 2012. 'Realist Randomised Controlled Trials: A New Approach to Evaluating Complex Public Health Interventions'. *Social Science & Medicine* 75 (12): 2299–2306. doi:10.1016/j.socscimed.2012.08.032.

Cartwright, Nancy, and Nick Cowen. 2014. *Making the Most of the Evidence in Education: A Guide for Working Out What Works .... Here and Now*. 2014-03. CHESS Working Paper. Durham University.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford ; New York: Oxford University Press.

Cartwright, Nancy, and Eileen Munro. 2010. 'The Limitations of Randomized Controlled Trials in Predicting Effectiveness: Limitations of RCTs for Predicting Effectiveness'. *Journal of Evaluation in Clinical Practice* 16 (2): 260–66. doi:10.1111/j.1365-2753.2010.01382.x.

Cookson, R. 2005. 'Evidence-Based Policy Making in Health Care: What It Is and What It Isn't'. *Journal of Health Services Research & Policy* 10 (2): 118–21. doi:10.1258/1355819053559083.

Cowen, Nick. 2012. *Rehabilitating Drug Policy*. London: Civitas. http://www.oecd.org/eco/labour/49421421.pdf.

Ellis, Sue. 2007. 'Policy and Research: Lessons from the Clackmannanshire Synthetic Phonics Initiative'. In *Approaching Difficulties in Literacy Development: Assessment, Pedagogy and Programmes*, 39–51. London: Sage. http://strathprints.strath.ac.uk/20642/1/strathprints020642.pdf.

Ellis, Sue, and Gemma Moss. 2014. 'Ethics, Education Policy and Research: The Phonics Question Reconsidered'. *British Educational Research Journal* 40 (2): 241–60. doi:10.1002/berj.3039.

Fixsen, Dean L., Sandra F. Naoom, Karen A. Blase, Robert M. Friedman, and Frances Wallace. 2005. 'Implemenation Research: A Synthesis'. http://centerforchildwelfare2.fmhi.usf.edu/kb/Implementation/Implementation%20Research%20-%20A%20Synthesis%20of%20Literature%20%20-%202005.pdf.

Gelman, Andrew, and Christian P. Robert. 2014. 'Revised Evidence for Statistical Standards'. Accessed April 28. http://www.stat.columbia.edu/~gelman/research/unpublished/Val_Johnson_Letter_3.pdf.

Goldacre, Ben. 2009. *Bad Science*. London: Harper Perennial.

———. 2013a. *Bad Pharma*. London: Fourth Estate.

———. 2013b. 'Building Evidence into Education'. http://dera.ioe.ac.uk/17530/1/ben%20goldacre%20paper.pdf.

Greenhalgh, T., J. Howick, N. Maskrey, and for the Evidence Based Medicine Renaissance Group. 2014. 'Evidence Based Medicine: A Movement in Crisis?'. *BMJ* 348 (jun13 4): g3725–g3725. doi:10.1136/bmj.g3725.

Harford, Tim. 2014. 'The Random Risks of Randomised Trials'. *Financial Times*, April 25, sec. Magazine. http://www.ft.com/cms/s/2/59bb202c-ca7b-11e3-8a31-00144feabdc0.html?utm_content=bufferbcdd0&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer#axzz2zyV8b8cw.

Healy, D. 2001. 'The Dilemmas Posed by New and Fashionable Treatments'. *Advances in Psychiatric Treatment* 7 (5): 322–27. doi:10.1192/apt.7.5.322.

Hochschild, Jennifer L. 2009. 'Conducting Intensive Interviews and Elite Interviews'. *Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. http://www.nsf.gov/sbe/ses/soc/ISSQR_workshop_rpt.pdf.

Holmes, Dave, Stuart J Murray, Amelie Perron, and Genevieve Rail. 2006. 'Deconstructing the Evidence-Based Discourse in Health Sciences: Truth, Power and Fascism'. *International Journal of Evidence-Based Healthcare* 4 (3): 180–86. doi:10.1111/j.1479-6988.2006.00041.x.

Honig, M. I., and C. Coburn. 2007. 'Evidence-Based Decision Making in School District Central Offices: Toward a Policy and Research Agenda'. *Educational Policy* 22 (4): 578–608. doi:10.1177/0895904807307067.

Houser, Janet, and Kathleen S. Oman, eds. 2011. *Evidence-Based Practice: An Implementation Guide for Healthcare Organizations*. Sudbury, MA: Jones & Bartlett Learning. http://sgh.org.sa/Portals/0/Articles/Evidence-based%20Practice%20-%20An%20Implementation%20Guide%20for%20Healthcare%20Organizations.pdf.

Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide. 2003. US Department of Education: Coalition for Evidence-Based Policy. http://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf.

Johnson, V. E. 2013. 'Revised Standards for Statistical Evidence'. *Proceedings of the National Academy of Sciences* 110 (48): 19313–17. doi:10.1073/pnas.1313476110.

Liverani, Marco, Benjamin Hawkins, and Justin O. Parkhurst. 2013. 'Political and Institutional Influences on the Use of Evidence in Public Health Policy. A Systematic Review'. Edited by Gemma Elizabeth Derrick. *PLoS ONE* 8 (10): e77404. doi:10.1371/journal.pone.0077404.

Nutley, Sandra, Huw Davies, and Isabel Walter. 2002. 'Evidence Based Policy and Practice: Cross Sector Lessons from the UK'. *ESRC UK Centre for Evidence Based policy and Practice: Working Paper* 9. http://79.125.112.176/sspp/departments/politicaleconomy/research/cep/pubs/papers/assets/wp9b.pdf.

Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman, and James Thomas. 2014. 'A Systematic Review of Barriers to and Facilitators of the Use of Evidence by Policymakers'. *BMC Health Services Research* 14 (1): 2. doi:10.1186/1472-6963-14-2.

Parsons, W. 2002. 'From Muddling Through to Muddling Up - Evidence Based Policy Making and the Modernisation of British Government'. *Public Policy and Administration* 17 (3): 43–60. doi:10.1177/095207670201700304.

Phillips, D. C. 2005. 'The Contested Nature of Empirical Educational Research (and Why Philosophy of Education Offers Little Help)'. *Journal of Philosophy of Education* 39 (4): 577–97. doi:10.1111/j.1467-9752.2005.00457.x.

Sharples, Jonathan. 2013. 'EVIDENCE FOR THE FRONTLINE'. Alliance for Useful Evidnce. http://www.alliance4usefulevidence.org/assets/EVIDENCE-FOR-THE-FRONTLINE-FINAL-5-June-2013.pdf.

Sheridan, Desmond J. 2013. *Medical Science in the 21st Century: Sunset or New Dawn?*. London: Imperial College Press.

Smith, George Davey, Shah Ebrahim, and Stephen Frankel. 2001. 'How Policy Informs the Evidence'. *British Medical Journal* 322 (7280): 184–85.

The National Strategies: A Review of Impact. 2010. 080270. London: Ofsted.

Wiggins, Meg, Helen Austerberry, and Harriet Ward. 2012. 'Implementing Evidence-Based Programmes in Children's Services: Key Issues for Success'. *Childhood Wellbeing Research Centre. Retrieved from: Www. Cabinet-Office. Gov. Uk/moderngov/policy/index. Htm.* http://holywellpark.info/media/wwwlboroacuk/content/ccfr/publications/dfe-rb245.pdf.

Young, Ken, Deborah Ashby, Annette Boaz, and Lesley Grayson. 2002. 'Social Science and the Evidence-Based Policy Movement'. *Social Policy and Society* 1 (03). doi:10.1017/S1474746402003068.

Ziliak, Stephen T., and Deirdre N. McCloskey. 2009. 'The Cult of Statistical Significance By Stephen T. Ziliak and Deirdre N. McCloskey Roosevelt University and University of Illinois-Chicago'. http://stephentziliak.com/doc/2009Zili