

What is the evidence on the best way to get evidence into use in education?

Stephen Gorard, Beng Huat See and Nadia Siddiqui  
Durham University Evidence Centre for Education  
s.a.c.gorard@durham.ac.uk

## Abstract

Over decades, there have been calls by concerned stakeholders to improve the quality of education research, and some progress has been made towards creating a more secure evidence base in some areas. More programmes and approaches that have a reasonable evidence base are also now being used in schools (but not in policy, and not necessarily because they have a reasonable evidence base). However, there has been no equivalent improvement in secure knowledge about how best to get that evidence into use, or even what difference it makes when such evidence is used. This paper looks at what little is already known about the different ways to get research evidence into use in education. It does so by summarising the results of a large-scale review of the literature. A total of 323 most relevant studies were found across all areas of public policy, and judged for quality and contribution. Very few (33) were of the appropriate design and quality needed to make robust causal claims about evidence-into-use, and even fewer of these concerned education. This means that despite over 20 years of modest improvement in research on what works in education policy and practice, the evidence on how best to deploy such evidence is still very weak.

We consider studies in terms of several issues, including whether they look at changes in user knowledge and behaviour or student outcomes, and how evidence is best modified before use. Providing access to raw research evidence or even slightly simplified evidence is not generally an effective way of getting it used, even if that evidence is presented to users by knowledge-brokers, in short courses or similar. What is more likely to work for both policy and practice, is engineering high quality evidence into a more usable format and presenting it actively or iteratively via a respected and trusted conduit, or through population measures such as legislation. Having the users actually do the research is another promising approach.

Expecting each individual study they fund to have impact is not the way forward as this may encourage widespread use of ineffective or even harmful interventions. Publicly-funded users, including policy-makers, should be required to use evidence-led programmes from such libraries, where they exist and are appropriate and relevant to their aims. Research funders should support these approaches, and help to build up libraries of successfully tested programmes. Researchers need to be scrupulous, looking at their new evidence in the context of what is already known and not looking to have ‘impact’ from single-studies. More and better research is needed on the best routes for evidence-into-use. However, the improvements required of all parties are as much ethical in nature as they are technical or scientific.

## Introduction

Evidence of different kinds has always been a part of education. For example, in England schools are given feedback to help improve after statutory inspection by Ofsted, and they also routinely look at and consider the implications of data on student context and outcomes (Schildkamp 2019).

However, there is increasing policy pressure to improve student outcomes at school by basing decisions about strategies and approaches on robust evidence generated by research (Hollands et al. 2019). Evidence derived from research is considered important in contributing to improvements, as agreed by most professionals working in public policy and practice (Davies 1999). High quality evidence can reportedly lead to important gains for children, the public, and society (Palmer 1999).

Therefore, in education policy and practice, there have rightly been increasing demands over 20 years from funders and governments for publicly-funded research to be of higher quality and to have real-world “impact” (Hillage et al. 1998, NRC 1999, Mosteller and Boruch 2002, Smith 2003). This increasing emphasis on generating more robust evidence of “What Works” is exemplified in the UK by some of the work funded by the Economic and Social Research Council (ESRC) Teaching and Learning Research Programme, ESRC Evidence Network, ESRC National Centre for Research Methods, ESRC Quantitative Methods Initiative, Nuffield Foundation Q-Step programme, and the Education Endowment Foundation (EEF) in England. These and other initiatives follow or shadow similar ones in the US such as the Institute of Education Sciences (IES), while other countries have since produced their own versions of many of these initiatives.

Their evidence often does not improve policy/practice directly but could instead save a considerable amount of money from being wasted on ineffective approaches, and so permit more effective use of limited time or funding. Knowing what does not work is therefore very valuable and too often ignored or downplayed. For example, the EEF in England, and the IES in US, have published the results of hundreds of evaluations so far, and the vast majority have not had positive results. This enables research users such as schools to make much more informed judgements about which interventions are likely to work in their context, and which will not.

The successes and limitations of this evidence-based movement are considered in a later section of the paper. Here it is sufficient to note that these initiatives have helped to improve the quality and range of research, and therefore the understanding of effective interventions to inform education policy and practice. A growth in experimental designs, coupled with natural experiments, better and more open official data, and data archiving, have all encouraged improvements in robust evaluations of proposed programmes in education.

The situation in terms of research evidence use in policy/practice is less clear. Helping to embed secure research findings into policy and practice has been a concern for at least as long as the concern about improved evidence itself (Lawrence 1990), and is worldwide (Koh et al. 2010, van de Goor et al. 2017). The UK ESRC want “Knowledge Exchange and Impact” embedded at the heart of all ESRC research, and they present awards to studies regarded as having outstanding impact. This engagement and impact is monitored via the website Research Fish long after the projects they funded are complete. The UK Research Excellence Framework (REF) for funding research in universities requires the submission of academic impact case studies, which are then assessed by panels as being of one of five levels of quality. Other funders and stakeholders similarly not only want high quality evidence to be generated but also for that robust evidence to be of value in real-life. In a largely applied field like education it makes sense for the taxpayers and charity-givers who fund the research to demand that it be useful and used.

In the US, the American Educational Research Association (AERA) offers advocacy training including on handling education research, to help citizens when dealing with their political representatives (<http://cqrcengage.com/aeraedresearch/AERAAdvocacyToolkit?0>). EdReports.org offers advice for practitioners on evaluating educational materials for use in the classroom. In the UK, there have been a number of practical infrastructure developments over recent years including ResearchED, the Chartered College of Teaching, and the Research Schools Network, which support teachers to adopt evidence-informed practices. There is increasing interest among educators in using more evidence-led approaches, and on-line discussions of its benefits, via personal accounts (<https://twitter.com/missllewis/status/1156857479076208641?s=12>). However, the evidence base for these approaches is unknown.

Evidence-based approaches are undoubtedly being used in schools, and probably more so than in the past, simply because we now have a better evidence base for the kinds of approaches that schools might want to use. This does not mean that schools are using them principally because they are evidence based. And in policy areas there are few signs that the use of evidence has increased anyway. We need to know more about how primary research evidence is best routed to users, so that they are more likely to use evidence-based approaches, and do so because they are evidence-based.

This paper discusses three problems with the agenda of this evidence-based movement, the quality of existing evidence to be used, and issues in summarising this evidence accurately. This is followed by a summary of a new review of the evidence on how best to get evidence into use. It discusses the methods used, the quality of the studies found, the barriers to evidence use, and how best to get evidence into use. The findings of this new review are summarised, and their implications **discussed**.

### **Why we urgently need better understanding of evidence-into-use**

Unfortunately, the apparent desire by all parties for research evidence to have a greater role, via real-life ‘impact’, has generated at least three problems. First, it is not at all clear that the research that has the greatest **real-life** impact is actually of the highest quality. For example, the REF impact case studies (see above) must be based on research judged by the relevant panel of reviewers to be of at least 2\* quality. This rating is from a scale of 0 (not really research of any value at all) to 4\* (world-leading). In REF2014 around 30% of all work was judged to be 4\* (world-leading), and 76% overall was judged to be 3\* (internationally excellent) or better. It is therefore astonishing that work considered to be among the lowest 24% in quality should be celebrated as having had any impact in real-life. In no way does this match the expressed ambition of having robust evidence-informed policy and practice. Similarly, studies have been given ESRC awards (prizes) for the quality of their impact simply because their results are in widespread use, even where the evidence does not warrant such impact, or in some cases where the impact actually contradicts the findings of the research itself. It is as though merely claiming to make a difference is considered ‘impact’ of the kind that is needed, whether it is evidence-led in reality or not.

This means that high quality research may be used inefficiently, weak research may have unwarranted impact, and researchers may be tempted to push for unwarranted impact, and to over-claim the impact that their work has had. There is clear evidence that many researchers do indeed distort and exaggerate their impact claims, feeling under pressure to do so by their funders’ emphasis on impact, and through a desire to do well in REF assessments (Chubb and Watermeyer 2017). Instead stakeholders and funders should insist that only high quality evidence should have impact, and that it only counts as impact if what is done by users is in strict accordance with that evidence. Of course, users will not simply read evidence and then act in some kind of automatic way. Evidence-led use only means that the users’ actions must be in *accord* with the evidence. This may mean they may do nothing, the evidence may modify what they do slightly, or it may completely transform an educational process. However, the ESRC giving an impact award to a study that showed that an educational intervention did not work, but was still widely used, and promoted by the researchers, is quite wrong. This illustrates the scale of the first problem, and that it affects even stakeholders who apparently want evidence to have real-life impact.

The second major problem is that the idea of single studies having any impact at all may be misconceived. There is little direct replication of studies in education. Nevertheless, when conducting a systematic review or similar it is clear that there are many studies reporting on any issue, and that both the quality and the substantive findings of these studies differ. It is important to focus on, or give more weight to, the most robust and trustworthy evidence. But it is also important to realise that a best evidence summary may contain contrary findings. A single study may be of high quality but **be** contradicted by a larger number of equally good studies. In this case, the ‘best bet’ is to base policy or practice on the overall picture from all good studies, using the contradictory study as a caution (because it may still be nearer the correct result). One study should probably not have much, if any, impact by itself, except as such a caution against intemperate action that goes against its findings. Awarding grades and prizes for the impact of specific pieces of research is therefore actually contrary to the very idea of evidence-informed education, and should cease. We also need considerable improvement in judging and agreeing what high quality research is for any context, and what the aggregated results of an entire field are (**as discussed further below**).

The third major problem is that progress in generating good evidence for real-life use has not been matched by an equivalent growth in knowledge of how such evidence is best handled in order to have appropriate impact (Powell et al. 2017). The claim that research has had impact is a clear causal one, and so ought to be evaluated to the same standard as the initial research itself. The DfE (Coldwell et al. 2017), EEF (2018), and the US Coalition for Evidence-based Policy, among others, have been seriously considering how to translate and implement research findings. However, at the moment, recommendations for the implementation of evidence are not themselves based on good evidence from evaluations of how to implement evidence. There is perhaps even less evidence on how to get evidence into use in education than in other fields (Wentworth et al. 2017). The situation has not changed much over 20 years (Smith 2000, Jones 2018). Bennet (2019) describes education today as being like 15<sup>th</sup> century medicine, when claims were treated as true if they sounded plausible, little or nothing was tested before use, and miracle cures were mainstream.

In combination, these three problems for evidence-into-use lead to wasted opportunities, and even harm for the education system. As a society, we need to use the best evidence available in the most effective way possible. This is a practical and conceptual issue with implications beyond education, and an ethical concern for the majority of the population who both fund education via their taxes and charity-giving, and use its services for themselves or their families. Large amounts of public money are still being spent around the world on education initiatives that have no basis in evidence, little chance of working, and are continued even when their ineffectiveness has been revealed, creating the possibility of harm (Carrier 2017). There is always a kind of opportunity cost, given that every unwarranted policy uses time, effort and resources that could have been used for a genuine improvement. All of education is damaged when persuasive but poor quality evidence has widespread influence, as happens routinely in the UK, with programmes like Brain Gym or Learning Styles still being used in classrooms and universities.

These problems, especially the third, are the focus of this paper, which reviews the evidence on how best to get good evidence into use. The paper summarises the methods we used for the review, and how we judge the quality of research, and then looks at the existing evidence on the different approaches to getting research evidence-into-use, the range of factors which influence the uptake of high quality research evidence by policy or practice, how these have been evaluated so far, and which are the most effective pathways for evidence-into-use in particular contexts.

## Methods used to conduct the review

The research question addressed in this paper is:

What is the existing evidence on the most effective routes for getting high quality research findings into use for education policy and practice?

The question was addressed via a widespread search and review, sorting reports into the rarer studies directly addressing the causal question (in bold in the text), and studies including previous systematic reviews, secondary data analyses, economic appraisal methods, implementation evidence, conceptual papers, and in-depth views from stakeholders and public consultations which are used as background and explanatory material.

We searched using EBSCOhost database, which covers British Educational Index (BEI), PsycINFO, ERIC, Applied Social Sciences Index and Abstracts (ASSIA) OpenDissertations, Education Abstracts, Education Administration Abstracts, PsycARTICLES, plus JSTOR, Web of Science databases, Ebook selection through First Search, and Google Scholar. We followed up citations in reports and reviews found in the searches, and added literature already known to the reviewers from previous work in the field. The search was designed to be as inclusive as possible to identify a wide range of both published and unpublished literature, such as dissertations/theses.

The search was intentionally broad and inclusive (with no geographical, disciplinary or date limiters). Following several iterations we found that individual search terms combined with syntax either found very little, or else included evidence on the primary research that was intended to be used, **or examples of where evidence was being used, rather than evidence on how best to get is used** (i.e. almost everything). This is because of the double use of terms like ‘evidence’ in searching for evidence on evidence-into-use. We found that the more successful search terms included the following combination of statements or synonyms of these **(such as engineer\* and use, translati\* and use, evidence-based and evidence-informed)**:

- Use of research in policy/practice
- Impact of using research evidence in policy/ practice
- Most effective way of getting policymakers/practitioners to use evidence

As may be envisaged, terms like ‘translation’ and ‘engineering’ found what was mostly redundant material, even with modifiers such as ‘educat\*’, or not ‘mechanical’. This all made the prioritisation of more relevant material a very skilled and labour intensive task. **But each search found thousands of potentially relevant research reports.** The search was conducted separately for each of the search phrases twice, once for policy and once for practice. Given the broad and open search, and the number of hits, we set the search **engines** to present the most relevant pieces first. We sorted **the** searches by relevance and stopped when the next five pages or 100 results after the last relevant articles revealed no more relevant items. **Other than using this limiting but reasonable heuristic, it is hard to envisage assessing over 3 million pieces in any other way.** We will have missed some even many reports, but are confident that this review is both the biggest ever done in this area, and unlikely to have transformed **its** conclusions through discovery of a few more studies.

The number of records identified from each of the databases is shown in Table **1**.

Table **1** – Number of research reports found in each database

Database	Number of hits
Ebscohost	31,291
Web of Science	1,717
Ovid MEDLINE	107
EBook selection through First Search	6,856
Jstor	263,242
Google Scholar	3,050,000
Total	3,353,213

Initial screening of titles and abstracts in order of relevance created a set of 582 distinct reports that were broadly about how to use evidence in real-life applications, based on a scan of titles, and abstracts where possible. More detailed screening removed 262 pieces as duplicates or substantive duplicates leaving only 320. Of these 22 were subsequently excluded as not relevant on reading the full piece, and a further 25 were found in the citations of previous systematic reviews. The final number of pieces was 323, **of which all were read by more than one of the authors where there was any doubt about its relevance or provenance, and of which 33 were assessed as at least moderately robust in terms of our causal research question. This robustness was assessed using the approach in the next section.** These, and a fuller description of the review methods, are covered **more fully** in Gorard et al. (2019). **Where available, this** paper focuses more on the subset of results concerning education.






### *Judging the quality of existing research*


It is crucial that only the most robust research evidence has any impact in real-life (Gorard et al. 2017). It could be, and probably already has been, disastrous for the education of individuals where incorrect or misleading evidence has had enhanced use in real-life (Parkhurst 2017). This is something that needs widespread agreement by all parties, based on a standard process for judging how trustworthy any body

of research is. Several generic procedures have been suggested to help assess the security of the findings from any study, or when synthesising the evidence on any topic. For example, the EPPI Centre (for systematic reviews) has developed a “Weight of Evidence” framework, and there are other alternatives with the same intended purpose in fields other than education. Our own approach was part of the development of the EEF padlock security rating, which has a similar purpose (while being more technical and involving less judgement). In our approach the quality of each research report, and therefore the trustworthiness of its findings, is judged based on the underlying research design, scale, missing data, quality and relevance of measurements, fidelity and validity (Gorard 2018).

The first step is to identify the information on each of these quality issues from the report of the research being assessed. If the report does not include important information, or is written in such a way that the reader has no way of understanding it, then the research must be rated as having low security, and considerable doubt must be cast on its findings. The ensuing five possible (but essentially arbitrary) levels of quality are summarised in Table 2, in terms of the first four key elements. Each study can be given a rating representing its lowest row description for any of the first four columns. The full procedure appears in Gorard et al. (2017). This ‘sieve’ has been reported as easy to use, and leads to consistent judgements between raters.

Table 2 – A ‘sieve’ to assist in the estimation of trustworthiness of any research study

Design	Scale	Completeness of data	Data quality	Rating
Strong design for research question	Large number of cases per comparison group	Minimal missing data, no evidence of impact on findings	Standardised, independent, pre-specified, accurate	4 
Good design for research question	Medium number of cases per comparison group	Some missing data, possible impact on findings	Standardised, independent, not pre-specified, some errors	3 
Weak design for research question	Small number of cases per comparison group	Moderate missing data, likely impact on findings	Not standardised, independent, or pre-specified, some errors	2 
Very weak design for research question	Very small number of cases per comparison group	High level of missing data, clear impact on findings	Weak measures, high level of error, too many outcomes	1 
No consideration of design	A trivial scale of study, or number is unclear	Huge amount of missing data, or not reported	Very weak measures, or accuracy not addressed	0 

The sieve approach underlies the judgements made in the rest of this paper about the quality of existing research on evidence-into-use.<sup>1</sup> In a later section of the paper, evaluations of how best to get evidence into use that are rated as 1, 2, 3 or 4  via this method are cited in bold. These all have at least some kind of counterfactual (or similar comparator), and are summarised in Tables 4 to 7 later in the paper. All other citations can be assumed to be of a lower quality of evidence, **in** being unable to provide a causal answer to the research question for this review, are descriptive or speculative, or may be high quality for another purpose but do not address the key question of how best to get evidence-into-use.

### *Types of evidence use*

This paper considers four distinct issues when identifying and classifying studies of the best ways to get robust research evidence into use. One is whether each study concerns policy such as legislation by central governments and funding arrangement by local authorities, or education practice such as how to organise a classroom and how to improve language teaching. However, too much can be made of this distinction between policy and practice. School and classroom evidence-based changes can influence larger policy, and policy can insist on the use of effective strategies or make them more or less feasible at school and classroom level (Slavin 2017).

Another dimension considered is whether the intention is to encourage users to be aware of and use research evidence in a general sense (such as by providing a toolkit or a menu of options), or whether the idea is to get them to use a specific piece of evidence for a specific purpose (such as by showing them how to implement **a new evidence-based practice**).

A third issue is about the intended outcomes of any attempt to encourage evidence use. Users may be found to have altered their attitude to, or knowledge of, research evidence. This is probably the easiest outcome to alter. Users may also be found to have altered their behaviour or practice in light of evidence. And these changes to behaviour may yield benefits for the end users of education policy/practice, such as improved test scores for students compared to those not being taught based on this evidence. This is almost certainly the hardest outcome to alter, but it is in the end a major purpose of the evidence-based movement.


The fourth issue relates to two further dimensions of getting evidence into use. Plain or raw evidence from research can simply be **read** by research users, as when policy-makers and their advisers are given access to research reports and articles. This is a cheap form of knowledge transfer, but it requires a skilful user to search, read, and summarise the evidence, and then to implement the changes. Alternatively, the raw evidence can be engineered into an artefact, or redrafted to another format, for easier use. But this may do injustice to the fuller evidence, and it makes the evidence harder for users to judge in terms of its underlying trustworthiness compared to similar artefacts that may not be as scrupulously evidence-based. We want to assess which approach is better.

Whether research findings are used as they are, or used to create an artefact that is meant to be easier for policy/practice, there can also be variation in how much activity goes into the attempted transfer of knowledge, by the researcher or developer. Findings can be simply made available, as with the popular Toolkit approach, or there can be more active engagement in explaining what the results are and why they should be trusted. The situation can be summarised as in Table 3. These nine snapshot cells form one of the ways in which the existing evidence is classified in this review.

Table 3 – Two dimensions of evidence-into-use

	Passive transfer	Engagement in transfer	(inter)Active transfer
Plain evidence	e.g. open access to journals	e.g. journal clubs	e.g. practitioner inquiry
Modified evidence	e.g. EEF Toolkit for practitioners	e.g. Think tanks	e.g. internships, research schools
Engineered evidence	e.g. lesson plans	e.g. hotlines, helpdesks	e.g. legislation for population measures

### The quality of existing research on evidence-into-use

Very little work on evidence-into-use has been found in our search that is anywhere near the level of quality and rigour that is currently required by users and funders when producing the evidence to be used in policy or practice in the first place (i.e. it is rated 0 ). Despite all of the progress that has been made over 20 years in improving research in education and beyond, almost none of this progress is

reflected in the evidence on **how to get** evidence-into-use. Only a small number of studies found in this review **(33)** have any kind of counterfactual to illustrate what would have happened if the approach (to evidence-into-use) suggested had not been used. Previous full reviews were found **that were** relevant to other policy areas such as health, but none **in education**.

**There** is a range of **work, such as** thought pieces on how evidence-use might work or be improved, or developing conceptual models for improving the uptake of evidence (Humes and Bryce 2001, Harper 2004, Zlotnik 2007, Green 2008, Lingard 2013, Brown 2014, Lubienski et al. 2014, Blase et al. 2015, Jordan and Cooper 2016, Farley-Ripple et al. 2018, Peterson 2018). Much is repetitive without apparent progress over many **years**.

**None** of the **ideas in these pieces** have been tested properly, and so none provide robust evidence of their effectiveness. The existing evidence on the relative ‘effectiveness’ of these approaches is largely limited to asking those involved in evidence use how it works, and what the barriers are, based on case studies, observational data, interviews and surveys, in passive research designs (Daugherty 2008, Honig and Coburn 2008, Sirat and Azman 2014, Urwick 2014, Briand-Lamarche et al. 2016, Ogden et al. 2016). There are also Toolkits and similar guidelines for evidence translation, or using evidence in policy-making, and ideas for training practitioners in evidence use (Winters and Echeverri 2012).

Some of the more promising previous work has tended to be conducted over a very short time, and at a very small-scale (Strambler and McKown 2013, Brown and Rogers 2015, Admiraal et al. 2016), even though change based on existing users adopting new ideas from up-to-date evidence often takes a long time to appear in more objective measures (Gardner et al. 2008, **Slavin et al. 2013**).

What is needed in order to address the research question for this paper are **more** robust evaluations of evidence into use, over a longer period, using a variety of routes and approaches. Some prior work has been done in this area, but the routes to, and mechanisms for, evidence use have rarely if ever been robustly evaluated (Langer et al. 2016). It is quite shocking how much research is reported about how to get evidence into use, but how little of it has been tested, and thus judged trustworthy.

### **The possible barriers to evidence use**

Neither policy-makers nor practitioners in education actually use much external research evidence (Mahoney 2013, Procter 2013, Galway and Sheppard 2015). Policy-makers and practitioners in education may say that they want and use good evidence (Penuel et al. 2017). However, few teachers incorporate evidence-led practices into their lesson plans or guidance for pupils. According to Wexler (2019), part of the fault lies with teacher training programmes in the US, because their lecturers are unaware of the best **available** evidence. Schools of Education have **usually** developed on a different track from the rest of academia, with little interest in research. Instead they teach as ‘gospel’ outdated and unwarranted theories. Much the same applies to CPD, based on non-evidenced practices. Teachers are often encouraged to adopt new **ideas** described by promoters as evidence-based, but **they** have no way of knowing if this is true. The label ‘evidence-based’ has itself become suspect. Personal experiences, accounts from other schools and teachers, and CPD not based on solid research evidence, are far bigger reported influences than research evidence, for practitioners (Nelson and Walker 2019, Walker et al. 2019).

Governments often start out at least espousing the use of evidence, but seem to become less interested and more dogmatic the longer they are in power (Perry et al. 2010). Policy-makers often select only a subset of the available evidence, especially where the research was conducted or paid for by their department, in order to help validate their political judgement. This is a process nicknamed ‘policy-based evidence **making**’.

**This** section outlines the possible barriers to use of evidence in education based on some of the conceptual and in-depth work about barriers and facilitators for getting evidence into use, as outlined



in the previous section. The stronger evidence, based on more suitable robust designs for the causal question of how to get evidence into use, appears in the later substantive sections of the paper.

### *Primary research quality*

An important barrier preventing the widespread use of high-quality evidence is the lack of high-quality evidence to use **in the first place**, caused chiefly by the lack of research competence among researchers themselves whose job it is to provide evidence in education. Twenty years ago, key UK stakeholders such as policy-makers, practitioner bodies, and senior researchers agreed reasonably well that education research is not good enough for real-life impact (Taylor and Gorard 2002). The biggest complaints were a shortage of ‘quantitative’ skills, limited attention paid to causal links, propensity for non-cumulative, small-scale and/or ‘flawed’ research. Following the initiatives outlined at the start of this paper, the situation is somewhat improved, but the problem of poor quality, largely unusable education research **remains (Hazell 2019)**. **In fact**, this kind of work may be more misleading because it tends to be taken more seriously by users than the more common small-scale, uninventive and sloppy journalistic, solely ‘qualitative’ work. The latter still forms the majority of UK education research, and has actually grown as a proportion of education research over the last 30 years (Galvez et al. 2019).

A **related** major barrier to coherent evidence use in real-life is the antagonistic stance of many researchers in education to the very idea that practical or scientific knowledge is possible in education (Biesta 2010, McKnight and Morgan 2019). There are criticisms of what counts as ‘evidence’, and suggestions that social reality is so complex and irreducibly rich that useful evidence of ‘what works’ is not even **possible (Marston and Watts 2003)**. A number of educational interventions validated by the What Works Clearinghouse have been reported as ineffective when rolled out into practice. This has led some commentators to conclude that using research evidence cannot improve education (Pogrow 2017), or even that evidence-led education is just not possible (Cain 2015).

It may well be difficult to uncover immediate benefits from evidence use in education. Even where an intervention has been *robustly* evaluated and found to be successful on many occasions in the past, it still may not ‘work’ when rolled out into practice or on all occasions or in all contexts. It is easy to find examples where a programme thought to be effective did not lead to improvement when used by others. This is to be expected in any scientific endeavour. Such programmes are not proven; they are only examples what has worked in the past. Similarly, it is possible to find examples of interventions that have no firm basis in research evidence but could improve outcomes (Vujcich et al. 2016). This happens perhaps through luck, individual teacher prowess, or other factors acting in parallel with the intervention. No approach has ever been absolutely proven to work or not to work. None of these points would be an argument for not using the approach that has the best chance of **success**.

**Another** area of weakness for researchers lies in their ability or willingness to communicate their findings to users (Miller 1999). The process of knowledge transfer needs more capacity-building work – perhaps especially in universities (Cooper et al. 2009, Perry et al. 2010). In order to influence policy, ‘policy entrepreneurs’ may need to be political fixers, networkers, and good storytellers (Alliance for Useful Evidence and Cabinet Office 2017). Universities or other research organisations may need to spend more on research translation, media and communication, and provide suitable incentives for academics to conduct the different kinds of public-facing work involved, and to create appropriate outputs for users. To some extent, this kind of improvement is already underway.

### *Skills and attitudes of the users*

Education is reputedly one of the most value-laden areas of public policy (ATL 2013). The attitudes and skills of users of research may be a particular problem in education, where so many people behave as ‘experts’ (Hemsley-Brown and Sharp 2003). Educators can often be sceptical about evidence, or treat it in a superficial way (Finnigan et al. 2013), depending upon their prior beliefs (Cook 2015). So, even high quality research will make no difference unless potential users are receptive to new knowledge (Brownson 2017).

There is growing evidence

**available** on how to improve education practice, which is increasingly simplified and presented as evidence-led materials and curricula, via interactive websites or Toolkits. **However**, some practitioners can feel constrained in acting in accordance with evidence because they lack the authority to change existing practice (Gerrish and Clayton 2004), or due to lack of resources (Felton 2014). **Sufficient** funds could **be made available** in schools to implement these evidence-led approaches, by not spending on approaches without a clear evidence base. A problem lies in the determination by heads and teachers to select an approach, implement it properly, and permit the time for it to take hold (Myracle 2019).

School leaders appear content to plan school improvement without referring to robust evidence (Graves and Moore 2018). **Some** are already explicitly stating that they will ignore evidence in their schools, with a backlash against evidence-informed approaches, the use of RCTs and the feeling of being told what to do in what they see as their own domains (Cain 2019). Fraser et al. (2018) suggest that school **principals tend** to rely on local informants and other educators when deciding which educational programmes are most useful to them. Real improvement might **therefore** depend primarily on users having a desire to use evidence-informed approaches, and being shown the benefits of doing so more clearly. In practice, teachers rely more on judgement, their own experience and perhaps classroom data than robust research. Teachers may see research evidence as inessential and merely ‘nice to have’ (Nelson and Walker 2019).

In order to use evidence to inform policy, policy-makers would need a range of skills and knowledge that they may not currently possess – such as the ability to source, evaluate and use evidence (Jackson et al. 2018). David Laws, a former Minister of Education in England, claimed that education policy-makers are not very good at dealing with evidence, their civil servants are not very good at providing them with secure evidence, and many policies rely largely on ‘hunches’ (The Guardian 2017). Policy-makers, as with all individuals, may be tempted to ignore evidence unfavourable to their chosen ideas. Even when users are suitably trained and can make good judgements about the quality and relevance of evidence for use in public policy issues in abstract examples, they are much worse at doing so when considering a specific real policy issue. In the latter context they tend to give greater weight to anecdotes and personal experiences (Crocco et al. 2017). Often, attitudes and behaviour must be changed at an individual, team, organisation and system level in order for a new evidence-informed approach to take hold (Harvey and Kitson 2015). Influencing a ministerial decision in the UK, **for example**, usually involves lining up the Minister, special advisors, analysts and civil servants to agree, and to put forward the same proposal.

Policy-makers also help to define what sort of evidence actually counts as robust. Evidence-based approaches such as trials of classroom programmes **may be** attractive to them, because any blame attaches to teachers for using (e.g. learning styles) or not using (e.g. phonics) programmes led by such RCT evidence. **But** the work of the EEF in England, for example, is largely silent on the structural features of the national education system (Cowen 2019). So it cannot pass judgement on or criticise things that policy-makers are mostly responsible for, and where these same policy-makers routinely ignore or resist evidence (Gorard 2018).

For both policy and practice **use** a major barrier is the time and effort needed to keep up with the necessarily changing evidence on any topic (Gerrish and Clayton 2004). The organisations that actually deliver education, such as schools, need more capacity to find, share, understand, and use research. Otherwise, even the best research will have little impact (Cooper et al. 2009). The University of Bristol (2017) found that teachers reported lack of time as a big barrier to engaging with evidence, but perhaps the biggest barrier for practice in education is that teachers are largely unaware of the availability of evidence (Sparks 2018).

In the US, the Education Commission of the States produced a primer for policy-makers on how to source suitable research, read it critically, and use it appropriately (Royal Society/British Academy

2018). Something similar could be tried for other settings such as the UK, and for other stakeholders such as practitioners, and such artefacts are emerging. One important step would be to make lawmakers state explicitly what any new policy is intended to achieve, so that success and failure can be evaluated properly (Hess and Little 2015). This probably involves changing the ways in which policy is made, so that policy-makers routinely assess the available evidence, ensure that it is embodied in their decisions, and continue to do this as new evidence comes to light (Sisk 1993). Above all, policy-makers and practitioners must want to use evidence, and to understand why doing so is both practically and ethically important. And the wider public must demand this as well. Again, this is an ethical issue more than a technical one.

#### *Other possible barriers to evidence use*

Other barriers include the lack of timeliness in the research process, and the multiplicity of other demands on users. Although most education research can be deemed relevant to a real-world issue by its producers, evidence is often still not seen as relevant by users, especially policy-makers, because it is not timely enough. UK funding mechanisms for research such as the REF and ESRC have the major advantage that they are largely independent of government, and so permit considerable academic freedom. However, ESRC funding can take up to one year to obtain, and REF funding occurs only every seven years or so. And academics have other priorities as well. This all means that other non-academic bodies have been set up to deliver quicker research findings to funders like the EEF, but can then come to rely on their funding support. This leads to a mutual dependency of a kind that is then an enemy of evidence-led improvement (Gorard et al. 2017).

Policy-makers value timely evidence relevant to specific time-limited policy issues (Thomas et al. 2015), and any mismatch in the timelines of the users and producers of research simply means that evidence is likely to be ignored. This is a common occurrence because ministers and even administrations generally have a short life-span, much shorter than the period needed for the proper implementation of education reform (Perry et al. 2010). There is also considerable mobility within the UK civil service. A former UK minister of education complained that civil servants rarely have enough knowledge to provide up-to-date evidence (The Guardian 2017). Building the capacity of policy-makers and others to use evidence can be regularly hindered by rapid staff turnover, and changes in administration (Waga et al. 2013).

Public policy is seldom based on solid evidence, and is largely the product of other influences and issues. These include the need for quick solutions or for secrecy, political promises and expediency (Alliance for Useful Evidence and Cabinet Office 2017), personalities, personal relationships, available funding, convenience, heavy marketing by vested interests (Hess and Little 2015, Slavin 2017), self-interest (Noyes and Adkins 2016), and ideology (Rutter 2012). Based on UK Prime Ministers' memoirs, Perry et al. (2010) suggest that politicians rely more on their instincts than evidence, distrusting much of the education establishment, and preferring to listen to a trusted few, or to their favourite think tanks. Based on local authorities' engagement with evidence in London, Al Hallami and Brown (2014) suggest that how evidence is treated varies according to the ideology of the majority political party in the local authority.

The power of the media has also increased, often forcing politicians' hands in a direction contrary to the evidence. However, media attention is not a simple route (a silver bullet) for researchers to influence policy-makers (Zimmerman 2004). And even newer forms of feedback to governments, such as social media, may be biasing policy-making away from more robust evidence.

#### **How robust evidence can be best summarised and presented?**

One of the dimensions introduced in Table 3 concerned how evidence is best presented, so that it is easy to use while remaining accurate and robust. This is a very topical issue in UK education research, addressed in this section based largely on the most robust studies of evidence-into-use. However, before

considering how best to present research findings to users, it is important to recognise that there are long-standing disputes between researchers about how to present their findings, even to each other, in a scientifically appropriate manner.

### *Single study findings*

For studies addressing causal questions, including randomised control trials and statistical modelling, many researchers have traditionally presented their findings in terms of significance tests such as ANOVA and t-tests. This practice is declining and hopefully will be completely discontinued soon. Social science studies very rarely meet the necessary requirements for using a significance test (such as full randomisation). The tests are being used incorrectly and are widely misunderstood by readers anyway. Above all, they do not provide an estimate of the scale or substantive importance of any result (Gorard et al. 2017).

Methods experts in medicine, psychology, sociology, and education, the APA, ASA and other bodies now advise against their use (Fidler et al. 2004). The American Journal of Public Health, Epidemiology, Basic and Applied Psychology, and numerous medical and ecological journals have banned their publication, as have most US medical journals (Starbuck 2016). For the IES, Lipsey et al. (2012) stated that significance testing and p-values give misleading results, and are ‘best avoided’. Advances are stalled and scientific discoveries threatened by the “ubiquitous misuse and tyranny” of the p-value (Stang et al. 2010, p.1). This can prevent the public as students or patients from obtaining the most appropriate treatment, and will therefore harm many of them at some stage in their lives. Of course, the same issues apply to all approaches, such as confidence intervals and power calculations, that are based on the same assumptions and using the same logic.

What is needed in presenting a research finding is some idea of the scale of any difference or pattern or trend (Lecoutre and Poitevineau 2011), the methodological limitations underlying it (Kline 2004), and a judgement about its substantive importance. This is usually achieved by publishing ‘effect’ sizes or standardised measures of difference, trend, or correlation, and the scale of the study, along with a judgement of the quality of the research (see above). Effect sizes do not rely on the same unlikely assumptions as significance tests – so that they can be used legitimately with non-random cases, where substantial data is missing, and so on. They emphasise the scale of any finding.

For a large, reasonable quality study, the substantive findings from using any of the current competing forms of effect size such as Cohen’s d or Glass’ delta will always be the same. It is confusing to have more than one, but which one is used makes no real difference in practice. Xiao et al. (2016) conducted a re-analysis of 22 findings from 17 randomised control trials. Where they used the same outcome variables in each calculation, the results for each trial were the same whichever version of the effect size was used. Xiao et al. (2016) went further and compared the 22 results based on a simple effect size with those based on linear regression, multi-level modelling, and Bayesian multi-level modelling. Again, for these well-structured trials (RCTs), and using the same outcome variables in each calculation, these different approaches made no difference to the substantive result. Although neither of these comparisons had their true implications highlighted by the authors of that paper, these are useful illustrations. It seems that a simple effect size gives the same results in a well-designed large study as any of the more complex methods of presenting an analysis. For that reason, effect sizes should always be used, based on the simplest ES suitable for the data involved, while more complex analyses are not needed for the results of good studies. The ensuing simpler presentation will be as high quality as the more complex one, and much easier for academics and others to understand. A basic minimum for a causal study should be to report the means/scores for each group and overall, their standard deviations, and the resulting effect sizes, for pre-intervention (if available), post-intervention, and gain/progress scores. There is little advantage in trying to convert effect sizes into something more every-day for better communication, like the months of progress used by EEF (Baird and Pane 2019).

### *Summarising multiple findings*

It would be unusual to envisage a single study providing a useful or definitive answer to a real-life question by itself. It is more appropriate to look at all of the relevant evidence **on any issue**. Again though, academics and researchers have argued among themselves about how this **aggregation** should be done fairly and accurately. There is agreement, among experts that any summary must be inclusive, and seek out both published and unpublished results to reduce publication bias, and that the methods used for searching and summarising must be explained. Effect sizes remain open to publication bias and abuse, just as significance tests were, despite being theoretically independent of sample **sizes** (Kuhberger et al. 2014). Perhaps this is because small samples are more likely to produce aberrant or extreme ‘effect’ sizes which, like significant results, are then more likely to be published (Slavin and Smith 2009). This is another reason why the quality of each study, including its scale, needs to be considered alongside the headline results.

Some reviewers will simply list and report the findings of all relevant studies, some will present the results thematically, and some will meta-analyse the findings. A meta-analysis is a particular kind of systematic review that combines the ‘effect’ sizes of all of the studies in the review to provide an aggregate ‘effect’ size, or an overall single answer to an effectiveness question. This is perfectly proper as long as full account is taken of the relative strength and clarity of each study, and the studies being aggregated are fully commensurable – same interventions, age range, outcomes and so on. The intention would not be to provide a precise measure of impact, but to help judge whether, on balance, an intervention or approach was likely to benefit or harm education. All studies for a meta-analysis also need to have key details reported in addition to the effect sizes, such as the attrition in each comparison group. Where the studies are not commensurable, or are included despite deficient reporting of details, or no account is taken of their quality, then the results of the meta-analysis would be very misleading. Sadly, the latter is the **very** common. Most meta-analyses take no account of study quality, and improperly aggregate very different kinds of studies and outcomes.

School decision-makers around the world have been increasingly influenced by hyper-analyses of prior evidence, that synthesise the results of many meta-analyses - such as those by Hattie (2008), described on its cover as revealing “teaching’s Holy **Grail**”, and similar attempts around the world. These are even more problematic because again they are combining very different kinds of studies taking no account of their quality, or of the quality of the studies making up each meta-**analysis. Commentators** are now realising and rehearsing their dangers (Wolgemuth et al. 2017).

**Even the best systematic** syntheses of evidence often have little impact in practice, without further translation and activity (Haines et al. 2004, Green et al. 2016). So how should research findings be converted into more usable forms?

### *Engineering evidence for use*

A number of commentators/studies suggest that research evidence is more likely to be used if its findings are simple or simplifiable (Noyes and Adkins 2016), or adapted to context (Bertram et al. 2018), perhaps as summaries, overviews or policy briefs (Chambers et al. 2011). Simpler and more defined **descriptions of interventions** appear to have a better likelihood of success (Langer et al. 2016). Or at least, presentations of the evidence need to be in user-friendly formats that help policy-makers in their decision-making (Pew Charitable Trusts 2014). **Rosenbaum et al. (2010)** found in two small trials (with only 15 cases in the smallest group) that providing a clear summary of findings can improve understanding and retrieval of evidence from a systematic review. However, this was **assessed as** part of an interactive workshop, and there is no evidence on whether practice changed as a result afterwards.

In the increasingly popular Toolkit approach, an inventory of approaches or programmes is created, and categorised in terms of the evidence for their effectiveness, cost and benefits. These are often then simply made available or publicised for users. This passive approach does lead to some engagement, as evidenced by downloads and similar metrics, so the idea is relatively popular. In fact, a version of the EEF Teaching and Learning Toolkit in England is now being developed in many countries (Collins 2018). But there is little or no evidence that Toolkits then influence practice, and none at all so far that

they lead to improved attainment (see below). It **might** be better if it was simply made impossible to use public funding for programmes that are not supported by robust **evidence**.

**It** is not clear that simply modifying research findings into easier formats leads to any better results, with a passive approach to transfer, according to the summary of systematic reviews by **Bero et al. (1998)**. In health care, merely disseminating best practice guidelines is insufficient to alter practice in most cases (Miller et al. 2006, Gonzales et al. 2012).

In three small trials (334 cases overall), **Buljan et al. (2017)** compared a scientific abstract, a plain language summary, and an infographic, as ways of presenting knowledge about the results of Cochrane systematic reviews in health. The audience included doctors (111 intervention, 60 control), university students (46 intervention, 18 control), and lay consumers (45 intervention, 54 control). All three groups preferred the infographic (or the summary if not given the infographic), and rated it better than the other formats for reading and user friendliness (effect sizes of around 0.5). However, the infographic was no better for actually imparting knowledge of the findings of the review. **Studies like this are important because they illustrate why the body of weaker evidence cited earlier in the paper, based on asking people what they prefer or thinks works, can be so misleading.**

**Di Noia et al. (2003)** compared offering health guidance materials for 188 professionals via pamphlets (55 cases), CD-ROM (64) or the internet (69). The CD-ROM and the internet were linked to greater improvements in self-efficacy and **self**-reported behavioural intentions, but the study did not provide actual changes in treatment as an outcome. In one Campbell review of 24 studies (including 20 RCTs), e-learning compared to no learning at all was effective in improving evidence-based health care knowledge and skills, but made no difference to attitudes and, most crucially, no difference to clinical behaviour, and was silent on patient outcomes (**Rohwer et al. 2017**). Across six trials (from eight studies in a review) of using summaries of evidence for decision-making, evidence summaries were deemed easier to understand than complete systematic reviews, but their ability to increase the use of systematic review evidence in policymaking was inconclusive (**Petkovic et al. 2010**).

Therefore, just presenting modified summaries of evidence to users, and expecting them to act upon it is very unlikely to work as a method of translation to **use**. A systematic review of methods of teaching new medical trainees about evidence-based medicine found nine RCTs and compared a number of modes of delivery (Ilic and Maloney 2014). These included lecture versus on-line delivery, lecture versus small group work, directed or self-directed **learning**, and different groupings by disciplines. **None of the approaches was better than any other.**

The Literacy Octopus trials evaluated the impact of providing teachers in a total of 13,323 schools with research summaries and evidence-based resources to improve teaching, with or without light touch support (**NFER 2017**). After two years, there was little or no increase in any of six measures of teachers' use of research, and no improvement in pupils' Key Stage 2 English scores compared with the control group. Simply disseminating research summaries and evidence-based resources to schools does not so far seem to be an effective way for research organisations to support schools in improving pupil outcomes. This is an important study both in terms of its scale, and because it is one of only a few that assessed the impact of trying to improve research uptake on student outcomes. Schools **will** need more help than this in transforming such materials into beneficial change in the classroom.

Table 4 provides a summary of the more suitable research **so far that addresses our** research question concerning how evidence is best modified before use. Most studies compare presenting users with a modified form of the research **findings and** the research findings, or another form of modification. There is no evidence that end-user outcomes (such as student test scores in education) were improved by use of evidence, and some evidence that they are not. There is some evidence on whether user behaviour/practice changed as a result, and the **weight of evidence** suggests that simply providing users with access to evidence makes no difference, whatever format is used. The stronger studies are mixed on whether providing users with access to suitably modified evidence makes a difference to their knowledge/attitude concerning use of research evidence. But overall, the format used for the delivery

of evidence does not seem to matter much. Other than access and simplification, something else is needed to get evidence into actual use.

Table 4 – Summary of high/medium quality evaluations of modifying evidence for use

Study	Research quality	Knowledge/ attitude	Behaviour/ practice	End-user outcomes
NFER 2017	4*	-	x	x
Bero et al. 1998	3*	x	-	-
Petkovic et al. 2010	3*	-	x	-
Rohwer et al. 2017	3*	/	x	-
Buljan et al. 2017	2*	x	-	-
Di Noia et al. 2003	1*	/	?	-
Rosenbaum et al. 2004	1*	/	-	-

Key: / means positive impact, x negative/neutral impact, ? inconclusive impact, - means no evidence

### How effective is active engagement in delivery

In a review of evidence from 18 systematic reviews of evidence-into-use in health practice, Bero et al. (1998) confirmed that the passive dissemination of research evidence is largely ineffective. There is little robust causal evidence on the types of intervention that actually encourage users to take account of systematic reviews in their decision making. Given that simply making evidence available in any format is not very effective, does a more active presentation of the evidence work better? One review, that also covered some RCTs reported separately in this paper, suggested that a more active approach was needed (Perrier et al. 2011).

A study by Taylor et al. (2004) compared a convenience sample of 145 health practitioners allocated to a half-day critical appraisal skills training workshop (73), or a wait list (72). The knowledge and ability to appraise the evidence in a systematic review were higher in the workshop group, but there was no difference in the attitude towards using evidence, and no evidence of subsequent changes in practice or behaviour. According to the review by Gülmezoglu et al. (2007), providing three interactive workshops on finding and using evidence from a reproductive health library led to more knowledge and use of the library, but made no difference to practice. Similarly, a trial involving an educational visit to help obstetricians and midwives select and use evidence from a Cochrane database found no difference in resultant practice between 12 obstetric units in the intervention group and 13 in the control (Wyatt et al. 1998).

Users of evidence for policy and practice are rarely trained specifically to handle plain evidence presented passively (Connell and Klem 2000). As foreshadowed when discussing barriers (above), it is clear that policy-makers at all levels from elected representatives to civil servants, when asked, report that training in the use of evidence helps them to make sense of it or prepare their own evidence briefs (Pappaioanou et al. 2003, Uneke et al. 2012, Uneke et al. 2015a, Uneke et al. 2015b). However, training may not be enough. Their capacity to use evidence might be improved through tailored workshops and extensive mentoring (Haynes et al. 2018). Policy-makers also require research skills, and this needs to be part of their job descriptions (Peirson et al. 2012). Despite this, training and dissemination workshops often have poor attendance after a fanfare start, especially for the most senior policy-makers (Dagenais et al. 2013). Policy-makers also report preferring having more interactive research mentors or knowledge brokers on standby (Rolle et al. 2011, Traynor et al. 2014). Committees set up for sharing policy evidence did not work well, as they were not supported by management (Hoeijmakers et al. 2013).

Previous work in health found no good evidence that educational outreach interventions were effective in changing the behaviour of practitioners (Siddiqi et al. 2005). In that systematic review, the authors found 44 studies that met their selection criteria, but many had weak designs, and the overall evidence

was inconclusive on outreach, use of mass media, and reminders. One-off workshops are not generally effective (Miller et al. 2006), and nor are mailings or short courses according to a review of 55 studies in the existing literature (Davis and Taylor-Vaisey 1997). No difference was found between the use of seminars and roundtables in encouraging the use of systematic review results in health policy (Dwan et al. 2015).

From a review including five RCTs and three interrupted time series analyses, Murthy et al. (2012) suggested that mass mailing a printed bulletin with evidence from systematic reviews about obstetric care may improve evidence-based practice. However this could only be for one key message, about a change that is simple to implement, and where the users are at least partly receptive already. Under other circumstances, or where more complex changes are demanded, this approach shows no promise.

An RCT involving the directors of 126 zoos (63 randomised to each group) concerning education about sun-screening suggests that tailoring dissemination materials with follow-up phone calls leads to no advantage over simply making generic material available (Lewis et al. 2005). A synthesis of evidence from two studies of promoting evidence-informed decisions in nursing (tertiary care), within a review of 30 studies, found that engagement in the form of educational meetings and having an evidence mentor made no difference to nurse engagement in evidence-led behaviours (Yost et al. 2015). Abdullah et al. (2014) reviewed 10 studies of mentoring and individual attention from more senior colleagues in health as aids in getting knowledge of evidence into use, but nine of these studies were multi-faceted with mentoring as only one component, making it impossible to draw a clear conclusion about the impact of mentoring itself. The remaining study suggested some improvements in nurse practitioner behaviour in relation to evidence, but said nothing about the impact for patients. An earlier similar review by Thompson et al. (2007) found relevant studies to be of low quality in general. The four studies synthesised were inconclusive overall. Two found that researcher-led educational meetings with nurse practitioners made no difference to evidence uptake. Two were deemed effective, but had additional components, such as a local opinion leader running the meeting.

So although users say they prefer these more active approaches, these may actually be no better than simply making evidence available, in terms of changing behaviour or improving outcomes. Overall, there is little reason to think that access to evidence will assist widespread use, even when there is active presentation such as training and workshops for users. Table 5 summarises these studies, and shows no work on end-user outcomes, a largely negative picture for changes to behaviour, and a rather mixed picture for changing user knowledge of evidence.

Table 5 – Summary of high/medium quality evaluations of active engagement in evidence-into-use

Study	Research quality	Knowledge/ attitudes	Behaviour/ practice	End-user outcomes
Siddiqi et al. 2005	3*	-	/	-
Murthy et al. 2012	3*	-	?	-
Taylor et al. 2004	2*	?	-	-
Abdullah et al. 2014	2*	-	?	-
Wyatt et al. 1998	2*	-	x	-
Yost et al. 2015	2*	-	x	-
Lewis et al. 2005	2*	x	x	-
Davis and Taylor-Vaisey 1997	1*	/	x	-
Gülmezoglu et al. 2007	1*	/	x	-
Thompson et al. 2007	1*	-	?	-

Key: / means positive impact, x negative/neutral impact, ? inconclusive impact, - means no evidence

#### *Conduits between research and use*

Perhaps what is needed is a more formal intermediary between research and its users.



A relatively common suggestion as a way forward is greater collaboration between researchers and policymakers (van de Arend 2016). This could be based on ‘learning communities’ for policy (or practice), whereby academics and researchers welcome officials and others into such communities, to discuss new evidence as it emerges, and how to incorporate it in their planning (Brown 2014). The Alliance for Useful Evidence and Cabinet Office (2017) give an example using research from six Asian and African countries. The Future Health Systems consortium worked on a set of key strategies for improving the uptake of evidence into policy, such as improving the technical capacity of policy-makers, better packaging of research findings, the use of social networks and the establishment of fora to assist in linking evidence with policy outcomes. There are many other examples of such communities of practice (Langlois et al. 2016).

Such relatively formal consortia are one example of a kind of conduit between research itself and its users. Others might include knowledge brokers (Dobbins et al. 2009a, Ward et al. 2009a, 2009b), think tanks, What Works Centres or clearing houses, and some charities. Such intermediaries may have an important function in translating research findings into more acceptable formats, and connecting researchers to potential users (Cooper et al. 2009, Edwards 2010). However, Dobbins et al. (2009b) found that having a knowledge broker of this kind had no clear impact on research use compared to tailored messaging and simple access to a registry of research. Their study was based on 108 health departments in Canada (72 randomised to intervention). Advocacy organisations can also pursue an agenda that is not clearly related to the overall evidence, taking a partial view but using their political and media expertise to influence policy outcomes (Malin and Lubienski 2015).

There is a danger that any intermediaries are subject to the same pressures as users, and so may distort or cherry-pick evidence to suit their user clients. It is not always clear whether think tanks, based on an ideology, are using evidence accurately or not (Reid 2016). Conduits must be firm, honest, independent and credible to all parties, which is rare. The conduit groups must have access both to new research and to internal government documents (Williams 2010). In policy, the World Health Organisation recommendations for mother and child nutrition may provide a good example of evidence presented both fairly and usefully (Benmarhnia et al. 2017). Several commentators commend the idea of the National Institute for Health and Care Excellence (NICE) for healthcare in the UK, and suggest an equivalent for other areas (Langer et al. 2016), including for education policy (Perry et al. 2010). A combination of such an evidence conduit and a learning community for education research has been proposed by the Royal Society/British Academy (2018) in the UK, and provisionally named the “Office for Educational Research”.

All of these ideas sound plausible. However, the very limited evidence on research offer no promise (Table 6).

Table 6 – Summary of high/medium quality evaluations of research conduits

Study	Research quality	Knowledge/ attitudes	Behaviour/ practice	End-user outcomes
Dobbins et al. 2009b	2*	-	x	-

Key: / means positive impact, x negative/neutral impact, ? inconclusive impact, - means no evidence

### Research champions

There is no evidence yet that just linking users and researchers in research projects, perhaps via user groups as encouraged by ESRC and others, is effective (McLean et al. 2018). Two-way secondments between government and research departments are feasible, and are being increasingly attempted. It is an approach that has potential but, like so many others, has not been tested directly, and is only considered via surveys and similar (Uneke et al. 2018). Research-practice partnerships have been appearing in the US, but again have not been evaluated properly yet (Coburn et al. 2013).

One of the approaches that schools have adopted to assist the use of evidence has been to identify research leads or champions, whose role is to engage with the evidence in any area and then relay this

in more digestible form to the rest of the school. Teachers report several challenges to this model, particularly staff turnover among research champions (Rose et al. 2017). There is a weak positive correlation between the level of engagement of teachers with research, and the attainment of their pupils. However, the relationship may well not be causal, or if it is causal then it may be inverse, perhaps because good teachers are anyway more likely to engage with research.

Recent pilot evaluations funded by the EEF in England, involving five self-selected schools, found little evidence that providing research evidence for teachers in this way makes any difference to their student attainment, at least in the short term (**Griggs et al. 2016**). The intervention was based on a designated research champion covering five schools (and 2,075 pupils), auditing schools' needs, and providing a variety of generic and targeted development activities. Providing research evidence in this way may not increase teachers' reported use of research. Similarly, **Speight et al. (2016)** found no evidence of impact on future teacher behaviour (only some changes in attitudes) from encouraging teachers in 10 self-selected primary schools (teaching 280 pupils) to use modified or summarised research evidence on issues such as **metacognition**. **Both** pilot evaluations suggested that such interventions need considerable support from school leaders (not just research leads).

**A study** by **Wiggins et al. (2019)** looked at the use of research leads to promote evidence use, in maths and English for year 10 and 11 pupils. **In this**, 40 schools were randomised to treatment or not, and around 9 to 13% of pupil scores are missing for any outcome. As found in other studies there was a high turnover of leads, which reduced continuity. At a very small cost per pupil, results in the treatment group improved by  $ES=0.09$  and  $0.04$  for the first and second years of the trial, in maths, and **by**  $0.05$  and  $0.03$  in English.

In Florida, 2011 to 2013, 60 schools were randomised to a "Using Data" intervention or not (**Cavalluzzo et al. 2014**). The treatment involved professional development and then technical assistance to help teachers use data to improve student learning maths in Grades 4 and 5. Teacher skills in using data, their reported use of data to make decisions, and their attitudes to data use, all improved. However, **overall** student performance in the annual state maths assessment did not improve relative to **the control**.

**Three** US school districts used composite rich evidence of teacher effectiveness in their HR decisions for six years, including hiring, firing, tenure decisions, professional development, promotions, and incentive payments (**Stecher et al. 2019**). Rates of effective teaching did not change, and there were no improvements in achievement or dropout rates.

Suffolk County Council in England joined up with the EEF to subsidise the cost of evidence-based interventions and training in their local schools, from 2016 to 2018. The **ensuing** evaluation by Gu et al. (2019) was light-touch with no counterfactual (but offered promise of improved outcomes for isolated schools and low-income pupils).

The **University of Bristol (2017)** evaluated the use of workshops or Research Learning Communities, in which experts discussed research evidence with evidence champions from schools. There was some indication that this led to more interest in research among teachers, but there was no overall improvement in reading results for 5,462 pupils in 119 schools (60 intervention, 59 randomised to control). It may, of course, take longer than two years of involvement for any impact to be seen. This study is important because it is one of only three found in this review that is a direct test of a method of implementing evidence for learner outcomes.

**Overall, the evidence on research champions or similar is stronger than in most approaches covered so far. This stronger evidence tends to show that the approach does not work for improving end-user outcomes, may be effective for changing behaviour, and looks very promising for changing attitudes (Table 7).**

**Table 7 – Summary of high/medium quality evaluations of delivering evidence-into-use**

Study	Research quality	Knowledge/ attitudes	Behaviour/ practice	End-user outcomes
Cavalluzo et al. 2014	4*	/	/	?
University of Bristol 2017	3*	/	-	x
Wiggins et al. 2019	3*	-	-	/
Speight et al. 2016	1*	/	x	-
Griggs et al. 2016	1*	-	-	x
Stecher et al. 2019	1*	-	x	x

Key: / means positive impact, x negative/neutral impact, ? inconclusive impact, - means no evidence

### *Ongoing interactive approaches*

There are interventions for getting evidence-into-use in the health sector that have been tested successfully using experimental approaches, and which are potentially replicable in an education setting. Individual coaching on evidence-informed approaches in specific areas, with individual feedback, is perhaps the most promising approach (Miller et al. 2006), although an evaluation of a dental hygiene intervention based in 20 cities, with 385 registered hygienists randomised to intervention and 366 to control, suggests that self-study can be almost as effective and, of course, cheaper to implement (Gordon et al. 2005). Some “training” interventions based on small group workshops and follow ups have shown that knowledge among health practitioners increased at post-test, but there were no differences in treatment behaviour or patients outcomes according to a trial of 148 medical practitioners, with 73 randomised to intervention (Forsetlund et al. 2003). Similarly using communities of practice with software follow up was linked via a RCT of 37 mental health practitioners (18 in intervention) to having greater knowledge of a mental health intervention, but not to reporting greater readiness to change practice, or reporting actual change (Barwick et al. 2009).

Interactive multi-component interventions can improve therapists’ and practitioners knowledge of evidence (Davis and Taylor-Vaisey 1997), and can then change their practice behaviour (but not their attitudes), compared to passive dissemination of evidence according to a review of 12 interventions by Menon et al (2009). The Translation Research in a Dental Setting (TRiADS) project set out to improve evidence-based practices among Scottish dentists using behaviour-change theoretical models from the psychological literature (Clarkson et al. 2010). This led to a large cluster RCT involving 2,566 dental practices (1,999 intervention and 567 control), which demonstrated that an audit and feedback intervention was effective at persuading dentists to undertake anti-biotic prescribing in line with the current evidence (Elouafkaoui et al. 2016). There is further evidence from a systematic review of health care interventions that audit and feedback is the most effective way to get evidence into use, at least for the short term and often combined with other approaches (Siddiqi et al. 2005), although some studies find the combination only moderately effective (Davis and Taylor-Vaisey 1997). In fact, there may be generally more evidence of effective approaches involving feedback, reminders and education for professionals such as dentists, rather than policy-makers or the public (Haines et al, 2004). However, an older RCT providing feedback and advice to 36 medical residents (18 per group) on ordering expensive tests for patients was inconclusive in terms of reduction of inappropriate testing for the treatment group (Ruangkanchanasetr 1993).

Table 8 provides a summary of the findings of the better studies concerning ongoing or iterative approaches to promoting evidence-use. There is no evidence on benefits for end-user outcomes. The stronger studies tend to show that attitudes and behaviour can be influenced in this way.

Table 8 – Summary of high/medium quality evaluations of ongoing interactive approaches

Study	Research quality	Knowledge/ attitudes	Behaviour/ practice	End-user outcomes
Elouafkaoui et al. 2016	4*	/	/	-
Siddiqi et al. 2005	3*	-	/	-
Menon et al 2009	3*	x	/	-

Forsetlund et al. 2003	2*	/	x	-
Gordon et al. 2005	2*	?	-	-
Davis and Taylor-Vaisey 1997	1*	/	x	-
Barwick et al. 2009	1*	/	x	-
Ruangkanchanasetr 1993	1*	-	x	-

Key: / means positive impact, x negative/neutral impact, ? inconclusive impact, - means no evidence

## User engagement in research

A more radical approach is for the users of research to be involved in generating the relevant evidence themselves. One promising but so far not evaluated way forward is the use of software to help users to commission their own systematic reviews (Campbell et al. 2011).

Teaching staff in England are reported as being less confident about engaging in research themselves, than they are about reading the research of others (Hammersley-Fletcher et al. 2015). Nevertheless, it has been argued that schools working alongside researchers would sharpen knowledge of what evidence means and how it can strengthen the work of schools.

See et al. (2016) evaluated an intervention in which a cluster of nine partnership primary schools agreed to look at a research journal article on enhanced classroom feedback, hold a series of cascading training events for all staff, and then conduct three action research cycles in just over one school year. The intention was to improve pupil attainment. Some teachers found it very hard to understand the primary evidence, and often implemented the findings in an inappropriate way. The academic outcomes and progress scores for pupils overall were no better than for five comparison schools in the local authority, and for all maintained schools nationally. Even such interactive transfer of primary evidence is therefore challenging for users. Similarly, Levin et al. (2011) found little difference from study groups or even school districts conducting research themselves in the take-up of evidence in schools' policy and practice in 11 districts in Canada. Teachers probably need clearer guidance, professional development, and modelling of effective strategies on the use of research evidence, to improve attainment. There needs to be some more engineered format to translate research evidence into practical guidance for teachers – which could be in the form of lesson plans or protocols - that are simply used by teachers without the need for them to understand or appreciate the underlying evidence base.

Sometimes, however, impact might be achieved by the potential users participating in research themselves (the ultimate interaction with research evidence), and not just accessing the evidence. For example, nurse-led research is gaining some attention as a critical pathway to practical and effective ways of improving patient outcomes (Curtis et al. 2017). Similarly, McLaughlin et al. (2000) reported that involving teachers and school administrators in research teams meant that the research was more targeted at their concerns, and the staff learnt more about research use, and were more likely to support use of the findings. Similar claims (and they really are only claims) have been made based on participatory research on aging (McWilliam 1997). In a small RCT run by Tranmer et al. (2002), 28 nurses, who worked in two clinical research groups on developing their own research protocol (18 high participation, 10 low participation), reported more use of research evidence in practice, than a comparator group of 207 nurses. An even higher gain was reported for the group that also discussed and critiqued the research literature among themselves.

A powerful equivalent in education could be randomised trials (i.e. not action research) set up and run by teachers themselves (Churches 2016). Both Siddiqui et al. (2015) and Gorard et al. (2016a) reported trials led by a set of unrelated schools, and then aggregated by the authors. Both trials were moderately successful. They showed that volunteer teachers were able to run robust evaluations with limited external guidance, and each intervention had a positive impact on student outcomes (effect sizes of around 0.25). The studies suggest that this approach to getting evidence into use by involving teachers in research is feasible. But they had no proper counterfactual for the evidence-into-use aspect – and

provide no indication that the two popular interventions would not have been used by these schools anyway even in the absence of the trial.

Table 9 summarises the strongest evidence in this area. It is still weak, and inconclusive. Users conducting research themselves is a promising idea that has not really been tested yet.

Table 9 – Summary of high/medium quality evaluations of user involvement in research

Study	Research quality	Knowledge/ attitudes	Behaviour/ practice	End-user outcomes
See et al. 2016	2*	-	-	x
Tranmer et al. 2002	1*	-	/	-

## Enforced use of evidence

Some commentators draw a distinction between evidence-informed policy/practice, referring to use that enhances the application of evidence with users’ involvement in the decision making process (Abbott et al. 2013), as described in the paper so far, and evidence-based use referring to a programme that facilitates the application of evidence without practitioners’ involvement in the decision making process (Melnyk 2007).

Successful UK REF case studies suggest that the under-pinning research having any real-life impact is usually invisible to teachers, and instead is best embedded in artefacts such as services and technologies (Cain and Allan 2017). These artefacts can either change attitudes or are simply used on trust, such as in pre-prepared lesson plans and resources. If transfer is to be passive then the intervention has to be so heavily engineered that its evidence-base is no longer clear to the user. An example would be clinical guidelines for nurses (Thomson et al. 2000).

The US Department of Education Investing in Innovation (i3) programme provided large grants for applicants wanting to improve attainment in their schools (<https://www.ed.gov/open/plan/investing-innovation-i3>). The largest grants required applicants to use only interventions rated as effective by the What Works Clearing House, from a short-list that had been robustly evaluated with success (Pogrow 2017). The US Every Student Succeeds Act requires that, to be eligible for school improvement funding, schools must use one of the top three categories of programmes – based on evidence-based interventions (California Department of Education n.d.).

Oddly, the equivalent scheme in England, the SSIF (Gov.UK 2017) gave schools large sums of money for improvement but had no requirement that the most promising approaches, based on scientific evidence, should be used. In fact, the Public Accounts Committee (2015) stated that the Department for Education (DfE) in England:

does not do enough to make sure... good practice is adopted in weaker schools... To date, the Department has supported schools to use the Pupil Premium [extra funding for schools taking poorer pupils] effectively primarily by funding the Education Endowment Foundation to carry out research into the evidence base for what works. However, it has done less to incentivise schools to use best practice and only recommends, rather than mandating, Pupil Premium Reviews for schools that do not use funding well. This is particularly worrying given concerns expressed to us that schools that perform poorly are less likely to seek out advice for themselves.

Nevertheless, the DfE now requires schools to account for their use of Pupil Premium funding, and most schools have websites describing the interventions that they have used, and the evidence-base for them. It is not clear that the evidence-bases are being judged appropriately, or that schools do not simply look for evidence after their decisions rather than basing the decision itself on the best evidence.

A similar idea to the compulsory use of Toolkits is also possible for policy-makers, and has been tried in Mississippi (Arinder 2016). All requests for new programmes would have to be screened for their supporting evidence including long-term cost-benefit. This could be the key role for legislative bodies, deciding on programmes not in terms of ideology but public interest and improvements for public services.

Kansagra and Farley (2012) draw a distinction between evidence-based interventions in health that are implemented at an individual level - like the prescription of Metformin for those diagnosed with diabetes, which is a treatment that requires initial and continuing actions by the patient - and those implemented at a ‘population’ level – such as the fluoridation of water to prevent dental caries. The latter have a lower cost per individual treated, and are more reliable than individual treatments. Another example might be using evidence from health science to plan urban spaces better to encourage health-enhancing exercise to reduce the incidence of chronic diseases (Giles-Corti et al. 2015). Or the imposition by international finance of, reportedly, evidence-based solutions on the government of developing countries (Williams 2016). An example from nursing might be the requirement for all clinical practice policies and procedures to be evidence-based (Oman et al. 2008). Such an approach has been enforced in health settings via the use of report cards (Valentine et al. 2014). It is likely that if evidence is to improve education successfully then more of such enforced or population-level measures need to be developed, validated, and implemented. However, these very promising approaches have not been clearly evaluated (Longjohn 2012).

Teachers may object to being made to follow heavily engineered evidence in this way. Where school chains like MATs (Multi Academy Trusts) specify not only the curriculum and teaching style but also provide a script, teachers describe feeling like robots (Roberts 2018). Of course, it is not at all clear in this example that the scripts in question are robustly evidence-based, but the same reaction would probably appear even if they were. The evidence used must be high quality, and it must be translated faithfully into appropriate educational artefacts.

One promising study is by Doabler et al. (2014) on the results of an evidence-led curriculum, in which it is assumed that teachers can follow the curriculum without being aware of the audit trail leading back to the evidence for its use. The study involved around 2,700 students in 46 schools in Oregon and Texas. A total of 379 observations were conducted in 129 classrooms (68 treatment and 61 control). The classroom teachers using the new curriculum delivered higher rates of practice opportunities for individuals and groups of students compared with comparison classroom teachers who implemented standard district mathematics instruction. No evidence is reported on student outcomes.

As shown in Table 10, only this last study meets minimum standards for a causal question on this approach, so while enforced/population use of evidence-led approaches is probably the most effective way to get evidence-into-use, considerably more research is needed before finally deciding that.

Table 10 – Summary of high/medium quality evaluations of engineered evidence use

Study	Research quality	Knowledge/ attitudes	Behaviour/ practice	End-user outcomes
Doabler et al. 2014	2*	-	/	-

## Summary

The review described here is not, and cannot be, like a standard systematic review on a precise question, and is nowhere described as such. It will be incomplete, in that there will be studies missed. The search will be hard to repeat as it was conducted over several months in a rapidly changing field, and without a simple search term because of the difficulties of distinguishing between any research evidence, evidence that evidence is being used, and evidence on how best to get evidence into use (only the latter concerned us here). There also no universally accepted term for the latter. There is multiplicity of ambiguous terms like ‘engineering’, ‘application’ or ‘translation’ for evidence use itself, but none for

the evidence on how to improve evidence use. Nevertheless, this will be the largest and most complete summary, and by some way, of the findings in this emerging area.

So far, a total of 33 relevant studies (or reviews that include such studies) have been found, with 48 distinct approaches evaluated within them. Many are not of very high quality but, unlike the merely descriptive work, all have at least a counterfactual design, or suitable comparator or similar. Much more robust evaluation work has been found on practice (across all fields) than on policy, and more in health than education or any other area. There is nothing specifically on education policy-making at all. This means that much of the improved evidence for education generated in the past 20 years could have been wasted or at least not used most effectively because, as a society, we have so little idea how to put evidence-into-use. Education policy is still not generally informed by evidence, but more so by politics, beliefs and the views of funders (Slavin 2017). This is true of policy more widely, even where decisions are clearly meant to be evidence-based (Dick et al. 2016). Users do not usually act in accordance with evidence (Epstein 2017). The findings of high quality substantive research have not become embedded in policy or practice, while too much lower quality evidence is still being promoted (Gorard 2018).

We need better studies of evidence-into-use in education, and policy-making more generally, especially looking at improvements in educational outcomes due to evidence use. There are currently too few, and the overall picture is unclear. It appears that even modifying evidence to be easier to use, and delivering it to users in an active way, leads to little or no improvement in outcomes – although it might improve practitioner knowledge and observed behaviour.

It is clearly possible to influence user knowledge and even their behaviour, but not by merely making research evidence available to them, even in a simplified format. Evidence needs to be modified at least to some extent, and presented actively and often iteratively. Even here, there are very few really good studies.

The most promising approach may be one where the use of evidence is enforced, perhaps by being built into the curriculum or by law. Such population measures, via legislation or structural change, remain promising ways of getting evidence-based approaches into use – but even such measures are currently too often not evidence-based. So even with these, the question remains – how is it possible to get the policy-maker or practitioner to recognise and use the best available evidence?

## Conclusions and implications

As with any attempt to change or improve behaviour, the role of money could be key in encouraging the use of evidence in policy/practice. Those who fund education research need to be responsible with the money they are entrusted with by tax-payers or charity-givers. The research they fund must be as high quality as possible, and the findings must be made as useful as possible. This is currently not happening. The most common approach used by funders to promote the use of evidence is to insist that users are linked to any project. Yet they are unable to provide any serious evidence that this approach is effective (McLean et al. 2018).

Clearly, the preparation and continuing development of teachers should have an evidence-led basis, which too often it currently does not. Here is another area where funding could play a role. Initial teacher training courses must be delivered, at least in part, by experts in education evidence. And the state should only fund them, and recognise their qualifications, if they are.

Policy and practice interventions should always be independently evaluated before reform takes place, instead of using rather haphazard pilots and phased rollout. Clear objectives must be pre-specified, and side effects taken into account. Incentives could be used, at least in the short term, to encourage users to rely more on evidence, and for the public to demand this (Miller et al. 2006, Rutter 2012). Public funds could be shifted towards only paying for programmes that have been demonstrated independently to have strong promise (Pew Charitable Trusts 2014, Hess and Little 2015). Good evidence of

effectiveness should be transparent in all policy and practice decisions about new programmes, and only those that offer a good return should be funded. A central repository of effective programmes should be built up by funders and others. In the same way that any area of research should start with a full review of existing evidence, so new results should also be placed clearly and coherently in the context of that prior evidence. Each new result should add to a kind of narrative “Bayesian” synthesis, considering how new research changes what we already think we know about in the repository, rather than seeking to have use and impact in its own right.

Those in charge of education reform must be responsible, and demand evidence-led policy and practice throughout the system. Programmes shown not to work, or where there has been no robust evaluation, should be actively discouraged, given that there is growing evidence of programmes that do seem to work. Researchers need to be equally responsible, and resist the clear demand for their evidence to be used, even if it is used incorrectly, by not conniving with invalid use just so that they can claim ‘impact’. These are all largely ethical issues, concerning the extent to which all of these stakeholders genuinely care about improving education.

### Acknowledgements

The work of this review was assisted by funding from the Faculty of Social Sciences and Health, Durham University, and by colleagues in the Durham University Evidence Centre for Education, especially Dr Naomi Griffin.

### References

- Abbott, P., McSherry, R. and Simmons, M. (2013) *Evidence-informed nursing: A guide for clinical nurses*, Routledge
- Abdullah, G., Rossy, D., Poeg, J., Davies, B., Higuchi, K., Sikora, L. and Stacey, D. (2014) Measuring the effectiveness of mentoring as a knowledge translation intervention, *Worldviews on Evidence-based Nursing*, 11, 5, 284-300
- Admiraal, W., Buijs, M., Claessens, W., Honing, T. and Karkdijk, J. (2016) Linking theory and practice: teacher research in history and geography classrooms, *Educational Action Research*, 1-16
- Al Hallami, M., and Brown, C. (2014) Scenarios of London Local Authorities' Engagement with Alliance for Useful Evidence and Cabinet Office (2017) What Works Centre Network, <http://www.alliance4usefulevidence.org/event/what-works-centres-network-open-day/>
- Alton-Lee, A. (2011) (Using) Evidence for Educational Improvement, *Cambridge Journal of Education*, 41, 3, 303-329
- Arinder, M. (2016) Bridging the divide between evidence and policy in public sector decision making, *Public Administration Review*, 76, 3, 394-398
- ATL (2013) *Using evidence to inform education policy and practice*, ATL
- Baird, M. and Pane, F. (2019) Translating standardized effects of education programs into more interpretable metrics, *Educational Researcher*, <https://journals.sagepub.com/doi/full/10.3102/0013189X19848729>
- Barwick, M., Peters, J. and Boydell, K. (2009) Getting to uptake: Do communities of practice support the implementation of evidence-based practice?, *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 18, 16-29
- Benmarhnia, T., Huang, J. and Jones, C. (2017) Lost in translation: piloting a novel framework to assess the challenges in translating scientific uncertainty, *International Journal of Health Policy & Management*, 6, 11, 649-660
- Bennet, T. (2019) The evidence revolution in education has begun, *teachwire*, <https://www.teachwire.net/news/the-evidence-revolution-in-education-has-begun>
- Bero, L., Grilli, R., Grimshaw, J., Harvey, E., Oxman, A. and Thomson, M. (1998) Closing the gap between research and practice: an overview of systematic reviews of interventions to promote



- the implementation of research findings: The Cochrane Effective Practice and Organization of Care Review Group, *BMJ (Clinical Research Ed.)*, 317, 7156, 465-468
- Bertram, M., Loncarevic, N., Radl-Karimi, C., Thogersen, M., Skovgaard, T. and Aro, A. (2018) Contextually tailored interventions can increase evidence-informed policymaking on health-enhancing physical activity: the experiences of two Danish municipalities, *Health Research Policy and Systems*, 16, 14
- Biesta, G. (2010) Why 'What Works' still won't work: From evidence-based education to value-Based education, *Studies in Philosophy and Education*, 29, 5, 491-503
- Blase, K., Fixsen, D. and Jackson, K. (2015) *Creating meaningful change in education: A cascading logic model*, Scaling-Up Brief 6, <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED576693&site=ehost-live>
- Briand-Lamarque, M., Pinard, R., Thériault, P. and Dagenais, C. (2016) Evaluation of the processes and outcomes of implementing a competency model to foster research knowledge utilization in education, *International Journal of Higher Education*, 5, 3, 168-181
- Brown, C. (2014) Advancing policy makers' expertise in evidence-use: A new approach to enhancing the role research can have in aiding educational policy development, *Journal of Educational Change*, 15, 19-36
- Brown, C. and Rogers, S. (2015) Measuring the effectiveness of knowledge creation activity as a means to facilitate evidence-informed practice: a study of early years' settings in Camden, London, *Evidence and Policy: A Journal of Research, Debate and Practice*, 11, 2, 189-207
- Brownson, R. (2017) *Dissemination and implementation research in health: translating science to practice*, Oxford: Oxford University Press
- Buljan, I., Malički, M., Wager, E., Puljak, L. Hren, D., Kellie, F., West, H., Alfirević, Z. and Marušić, A. (2017) No difference in knowledge obtained from infographic or plain language summary of a Cochrane systematic review: three randomized controlled trials, *Journal of Clinical Epidemiology*, <http://dx.doi.org/10.1016/j.jclinepi.2017.12.003>
- Cain, T. (2015) Teachers' engagement with published research: addressing the knowledge problem, *The Curriculum Journal*, 26, 3, 488-509
- Cain, T. (2019) Don't be a research fundamentalist, *TES*, 25/1/19, <https://www.tes.com/news/tes-magazine/tes-magazine/dont-be-a-research-fundamentalist>
- Cain, T. and Allan, D. (2017) The invisible impact of educational research, *Oxford Review of Education*, 10.1080/03054985.2017.1316252
- California Department of Education (n.d.) *Evidence-based interventions under the ESSA*, <https://www.cde.ca.gov/re/es/evidence.asp>
- Campbell, D., Donald, B., Moore, G. and Frew, D. (2011) Evidence Check: knowledge brokering to commission research reviews for policy, *Evidence and Policy*, 7, 1, 97-107
- Carrier, N. (2017) How educational ideas catch on: the promotion of popular education innovations and the role of evidence, *Educational Research*, 59, 2
- Cavalluzzo, L., Geraghty, T., Steele, J., Holian, L., Jenkins, F., Alexander, J. and Yamasaki, K. (2014) "Using Data" to inform decisions: Results from a randomized experiment of program efficacy, <https://eric.ed.gov/?id=ED555557>
- Chambers, D., Wilson, P., Thompson, C., Hanbury, A., Farley, K. and Light, K. (2011) Maximizing the impact of systematic reviews in health care decision making: a systematic scoping review of knowledge-translation resources, *The Milbank Quarterly*, 89, 1, 131-156
- Chubb, J. and Watermeyer, R. (2017) Artifice or integrity in the marketplace of research impact?, *Studies in Higher Education*, 14, 2, 2360-2372
- Churches, R. (2016) *Closing the gap: test and learn* (Research report DFE-RR500b), Nottingham: National College for Teaching and Leadership
- Clarkson J., Ramsay, C., Eccles, M., Eldridge, S., Grimshaw, J. et al. (2010) The translation research in a dental setting (TRiADS) programme protocol, *Implementation Science*, 5, 57, 10.1186/1748-5908-5-57
- Coburn, C., Penuel, W. and Geil, K. (2013) *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*, William T. Grant Foundation, New York, NY

- Coldwell, M., Greany, T., Higgins, S., Brown, C., Maxwell, B., Stiell, B., Stoll, L., Willis, B. and Burns, H. (2017) *Evidence-informed teaching: an evaluation of progress* (DFE-RR-696) London: DfE
- Collins, K. (2018) *Going global*, [https://educationendowmentfoundation.org.uk/news/going-global-scaling-up-evidence-to-the-international/?mc\\_cid=09c0b25e5b&mc\\_eid=ba4641e5ba](https://educationendowmentfoundation.org.uk/news/going-global-scaling-up-evidence-to-the-international/?mc_cid=09c0b25e5b&mc_eid=ba4641e5ba)
- Connell, J. and Klem, A. (2000) You can get there from here: Using a theory of change approach to plan urban education reform, *Journal of Educational and Psychological Consultation*, 11, 1, 93-120
- Cook, C. (2015) A supportive beliefs intervention to facilitate the implementation of evidence-based practices within a multi-tiered system of supports, *School Mental Health*, 7, 1, 49-60
- Cooper, A., Levin, B. and Campbell, C. (2009) The growing (but still limited) importance of evidence in education policy and practice, *J Educ Change*, 10, 159–171
- Cowen, N. (2019) For whom does ‘what works’ work? The political economy of evidence-based education, *Educational Research and Evaluation*, [http://www.academia.edu/37248352/For\\_whom\\_does\\_what\\_works\\_work\\_The\\_political\\_economy\\_of\\_evidence-based\\_education](http://www.academia.edu/37248352/For_whom_does_what_works_work_The_political_economy_of_evidence-based_education)
- Crocco, M., Halvorsen, A., Jacobsen, R. and Segall, A. (2017) Teaching with evidence, *Phi Delta Kappan*, 98, 7, 67-71
- Curtis, K., Fry, M., Shaban R. and Considine, J. (2017) Translating research findings to clinical nursing practice, *Journal of Clinical Nursing*, 26, 862–872, doi: 10.1111/jocn.13586
- Dagenais, C., Queuille, L. and Ridde, V. (2013) Evaluation of a knowledge transfer strategy from a user fee exemption program for vulnerable populations in Burkina Faso, *Global Health Promotion*, 20, 1 suppl, 70-79
- Daugherty, R. (2008) Reviewing national curriculum assessment in Wales: how can evidence inform the development of policy?, *Cambridge Journal of Education*, 38, 1, 73-87
- Davies, P. (1999) What is evidence-based education?, *British Journal of Educational Studies*, 47, 2, 108-121
- Davis, D. and Taylor-Vaisey, A. (1997) Translating guidelines into practice: a systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines, *Canadian Medical Association Journal*, 157, 4, 408-416
- Di Noia, J., Schwinn, T., Dastur, Z. and Schinke, S. (2003) The relative efficacy of pamphlets, CD-ROM, and the Internet for disseminating adolescent drug abuse prevention programs: an exploratory study, *Preventative Medicine*, 37, 646-653
- Dick, A., Rich, W. and Waters, T. (2016) *Prison vocational education and policy in the United States*, New York: Palgrave Macmillan
- Doabler, C. , Nelson, N., Kosty, D., Fien, H., Baker, S., Smolkowski, K., and Clark, B. (2014) Examining teachers’ use of evidence-based practices during core mathematics instruction, *Assessment for Effective Intervention*, 39, 2, 99 –11
- Dobbins M., Hanna S., Ciliska D., Manske S., Cameron R., Mercer SL., O'Mara L., DeCorby K. and Robeson P. (2009b) A randomized controlled trial evaluating the impact of knowledge translation and exchange strategies, *Implementation Science*, 23, 4, 61
- Dobbins, M., Robeson, P., Ciliska, D., Hanna, S., Cameron, R., O'Mara, L., ... Mercer, S. (2009a) A description of a knowledge broker role implemented as part of a randomized controlled trial evaluating three knowledge translation strategies, *Implementation Science*, 4, 1, 23
- Dwan, K., McInnes, P. and Mazumdar, S. (2015) Measuring the success of facilitated engagement between knowledge producers and users: a validated scale, *Evidence and Policy*, 11, 2, 239-252
- Edwards, M. (2010) Making research more relevant to policy evidence and suggestions, in Bammer, G., Michaux, A. and Sanson, A. (Eds.) *Bridging the 'Know-Do' Gap* (pp. 55-64), ANU Press
- EEF (2018) *IPEELL: using self-regulation to improve writing* (re-grant), [https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/ipeell?mc\\_cid=febc011cb7&mc\\_eid=ba4641e5ba](https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/ipeell?mc_cid=febc011cb7&mc_eid=ba4641e5ba)
- Elouafkaoui, P., Young L., Newlands R., Duncan E., Elders A., Clarkson J. et al. (2016) An audit and feedback intervention for reducing antibiotic prescribing in general dental practice, *PLOS Med* 13, 8

- Epstein, D. (2017) *When evidence says no, but doctors say yes*, ProPublica, <https://www.theatlantic.com/health/archive/2017/02/when-evidence-says-no-but-doctors-say-yes/517368/>
- Epstein, D. (2017) *When evidence says no, but doctors say yes*, ProPublica, <https://www.theatlantic.com/health/archive/2017/02/when-evidence-says-no-but-doctors-say-yes/517368/>
- Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. *Educational Researcher*, 0013189X18761042.
- Felton, M. (2014) *A mixed-methods inquiry into science teachers' perceptions of the effects of professional development experiences on implementation of research-based instructional practices*, University of New Orleans thesis, <https://scholarworks.uno.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=2838&context=td>
- Fidler, F., Thomson, N., Cumming, G., Finch, S. and Leeman, J. (2004) Editors Can Lead Researchers to Confidence Intervals, but Can't Make Them Think: Statistical Reform Lessons From Medicine, *Psychological Science*, 15, 2, 119-126
- Finnigan, K., Daly, A. and Che, J. (2013) Systemwide reform in districts under pressure The role of social networks in defining, acquiring, using, and diffusing research evidence, *Journal of Educational Administration*, 51, 4, 476-497
- Forsetlund, L., Talseth, K., Bradley, P., Nordheim, L. and Bjorndal, A. (2003) Many a slip between cup and lip. Process evaluation of a program to promote and support evidence-based public health practice, *Evaluation Review*, 27, 179-209
- Fraser, C., Herman, J., Elgie, S. and Childs, R. (2018) How school leaders search for and use evidence, Galvez, S., Heiberger, R. and McFarland, D. (2019) Paradigm Wars Revisited: A Cartography of Graduate Research in the Field of Education (1980–2010), *American Educational Research Journal*, <https://doi.org/10.3102/0002831219860511>
- Galway, G. and Sheppard, B. (2015) Research and evidence in education decision-making: A comparison of results from two pan-Canadian studies, *Education Policy Analysis Archives* 23, 108/109, 1-37
- Gardner, J., Holmes, B. and Leitch, R. (2008) Where there is smoke, there is (the potential for) fire: Soft indicators of research and policy impact, *Cambridge Journal of Education*, 38, 1, 89-104
- Gerrish, K. and Clayton, J. (2004) Promoting evidence-based practice: an organizational approach, *Journal of Nursing Management*, 12, 2, 114-123
- Giles-Corti, B., Sallis, J. F., Sugiyama, T., Frank, L., Lowe, M. and Owen, N. (2015) Translating active living research into policy and practice: one important pathway to chronic disease prevention, *Journal of Public Health Policy*, 36, 2, 231-243
- Gonzales, R., Handley, M., Ackerman, S. and O'Sullivan, P. (2012) Increasing the translation of evidence into practice, policy, and public health improvements, *Academic Medicine*, 87, 3, 271-278
- Gonzales, R., Handley, M., Ackerman, S. and O'Sullivan, P. (2012) Increasing the translation of evidence into practice, policy, and public health improvements, *Academic Medicine*, 87, 3, 271-278
- Gorard, S. (2018) *Education policy: Evidence of equity and effectiveness*, Bristol: Policy Press
- Gorard, S., Griffin, N., See, BH and Siddiqui, N. (2019) *How can we get educators to use research evidence?*, US: Lulu Press, 193 pages
- Gorard, S., See, BH and Siddiqui, N. (2017) *The trials of evidence-based education*, London: Routledge
- Gorard, S., Siddiqui, N. and See, BH (2016) An evaluation of Fresh Start as a catch-up intervention: And whether teachers can conduct trials, *Educational Studies*, 42, 1, 98-113
- Gordon, J., Andrews, J., Lichtenstein, E., Severson, H. and Akers, L. (2005) Disseminating a smokeless tobacco cessation intervention model to dental hygienists: A randomized comparison of personalized instruction and self-study methods, *Health Psychology*, 24, 5, 447-455
- Gov.UK (2017) *Strategic School Improvement Fund*, <https://www.gov.uk/guidance/strategic-school-improvement-fund>

- Graves, S. and Moore, A. (2018) How do you know what works, works for you? An investigation into the attitudes of senior leaders to using research evidence to inform teaching and learning in schools, *School Leadership and Management*, 38, 3, 259-277
- Green, C., Taylor, C., Buckley, S. and Hean, S. (2016) Beyond synthesis: Augmenting systematic review procedures with practical principles to optimise impact and uptake in Educational Policy and Practice, *International Journal of Research and Method in Education*, 39, 3, 329-344
- Green, L. (2008) Making research relevant: if it is an evidence-based practice, where's the practice-based evidence?, *Family Practice*, 25, suppl\_1, i20-i24
- Griggs, J., Speight, S. and Javiera, C. (2016) *Ashford Teaching Alliance Research Champion: Evaluation report and executive summary*, London: Education Endowment Foundation
- Gu, Q., Hodgen, J., Adkins, M. and Armstrong, P. (2019) *Incentivising schools to take up evidence-based practice to improve teaching and learning: Evidence from the evaluation of the Suffolk Challenge Fund*, Educational Endowment Foundation
- Gülmezoglu, A., Langer, A., Piaggio, G., Lumbiganon, P., Villar, J. and Grimshaw, J. (2007) Cluster randomized trial of an active, multifaceted information dissemination intervention based on The WHO Reproductive health library to change obstetric practices, *BJOG*, 114, 1, 16-23
- Haines, A. Kuruvilla, S. and Borchert, M. (2004) Bridging the implementation gap between knowledge and action for health, *Bulletin of the World Health Organization*, 82(10)
- Hammersley-Fletcher, L., Lewin, C., Davies, C., Duggan, J., Rowley, H. and Spink, E. (2015) *Evidence-based teaching: advancing capability and capacity for enquiry in schools: interim report*, Nottingham: National College for Teaching and Leadership
- Harper, H. (2004) *The role of research in policy development : school sex education policy in Scotland since devolution*, (Electronic Thesis or Dissertation), <http://search.ebscohost.com/login.aspx?direct=true&db=ddu&AN=B073948971608C95&site=ehost-live> Available from EBSCOhost ddu database.
- Harvey, G. and Kitson, A. (2015) Translating evidence into healthcare policy and practice: single versus multi-faceted implementation strategies—is there a simple answer to a complex question?, *International Journal of Health Policy and Management*, 4, 3, 123
- Hattie, J. (2008) *Visible learning*, London: Routledge
- Haynes, A., Rowbotham, S., Redman, S., Brennan, S., Williamson, A. and Moore, G. (2018) What can we learn from interventions that aim to increase policy-makers' capacity to use research? A realist scoping review, *Health Research Policy and Systems*, 16, 31
- Hazell, W. (2019) Education research 'has a problem and must get its house in order', *TES*, 5/2/19, <https://www.tes.com/news/exclusive-education-research-has-problem-and-must-get-its-house-order>
- Hemsley-Brown, J. and Sharp, C. (2003) The use of research to improve professional practice: A systematic review of the literature, *Oxford Review of Education*, 29, 4, 449-471
- Hess, F. and Little, B. (2015) "Moneyball" for education, American Enterprise Institute
- Hillage, J., Pearson, R., Anderson, A., Tamkin, P. (1998) *Excellence on research in schools*, Sudbury: DfEE
- Hoeijmakers, M., Harting, J. and Jansen, M. (2013) Academic Collaborative Centre Limburg: a platform for knowledge transfer and exchange in public health policy, research and practice? *Health Policy*, 111, 2, 175-183
- Hollands, F., Pan, Y. and Escueta, M. (2019) What is the potential for applying cost-utility analysis to facilitate evidence-based decision making in schools?, *Educational Researcher*,
- Honig, M. and Coburn, C. (2008) Evidence-based decision making in School District Central Offices: Toward a policy and research agenda, *Educational Policy*, 22, 4, 578-608
- Humes, W., and Bryce, T. (2001) Scholarship, research and the evidential basis of policy development in education, *British Journal of Educational Studies*, 49, 3, 329-352
- Ilic D. and Maloney S. (2014) Methods of teaching medical trainees evidence-based medicine: a systematic review, *Medical Education*, 48, 2, 124-135
- Jackson E., Purnawati E., Shaxson L. (2018) How and when do policymakers use evidence? Taking politics into account, in Pellini A., Prasetyamartati B., Nugroho K., Jackson E., Carden F. (eds) *Knowledge, Politics and Policymaking in Indonesia*, Springer: Singapore

- Jones, G. (2018) Where's the evidence for evidence-based practice improving pupil outcomes?, <https://www.garyrjones.com/blog/>
- Jordan, E. and Cooper, P. (2016) *Building bridges*, Child Trends, Research Brief 2, <https://www.childtrends.org/wp-content/uploads/2016/10/2016-56.1BuildingBridgesSharingWithPolicymakers.pdf>
- Kansagra, S. and Farley, T. (2012) Translating research into evidence-based practice, *American Journal of Public Health*, 102, 8, Letters
- Kline, R. (2004) *Beyond significance testing: Reforming data analysis methods in behavioral research*, Washington, DC: American Psychological Association
- Koh, H., Oppenheimer, S., Massin-Short, S., Emmons, K., Geller, A. and Viswanath, K. (2010) Translating research evidence into practice to reduce health disparities: a social determinants approach, *American Journal of Public Health*, 100, S1, S72-S80
- Kuhberger, A., Fritz, A. and Schemdl, T. (2014) Publication bias in psychology, *PLOSone*, doi: 10.1371/journal.pone.0105825
- Langer, L., Tripney, J. and Gough, D. (2016) *The science of using science: Researching the use of research evidence in decision-making*, London: EPPI-Centre, UCL Institute of Education
- Langlois, E., Montekio, V., Young, T., Song, K., Alcalde-Rabanal, J. and Tran, N. (2016) Enhancing evidence informed policymaking in complex health systems: lessons from multi-site collaborative approaches, *Health Research Policy and Systems*, 14, 20
- Lawrence, R. (1990) Diffusion of the US Preventive Services Task Force recommendations into practice, *Journal of General Internal Medicine*, 5, 2, S99-S103
- Lecoutre, B. and Poitevineau, J. (2011) *The Significance Test Controversy and the Bayesian Alternative*, <http://statprob.com/encyclopedia/SignificanceTestControversyAndTheBayesianAlternative.html>
- Levin, B., Cooper, A., Arjomand, S. and Thompson, K. (2011) Research use and its impact in secondary schools (978-1-896660-50-9), <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED535436&site=ehost-live>
- Lewis, E., Mayer, J., Slymen, D., Belch, G., Engelberg, M., Walker, K., . . . Elder, J. (2005) Disseminating a sun safety program to zoological parks: The effects of tailoring, *Health Psychology*, 24, 5, 456-462
- Lingard, B. (2013) The Impact of research on education policy in an era of evidence-based policy, *Critical Studies in Education*, 54, 2, 113-131
- Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*, Washington DC: Institute of Education Sciences
- Longjohn, M. (2012) Translating research into evidence-based practice, *American Journal of Public Health*, Letters, 102, 8
- Lubienski, C., Scott, J. and DeBray, E. (2014) The politics of research production, promotion, and utilization in educational policy, *Educational Policy*, 28, pp.131-144
- Mahoney, S. (2013) *Educational Research in the United States: A Survey of Pre-K-12 Teachers' Perceptions Regarding the Purpose, Conceptions, Use, Impact, and Dissemination* (Doctoral dissertation, Arizona State University)
- Malin, J. and Lubienski, C. (2015) Educational expertise, advocacy, and media influence, *Education Policy Analysis Archives*, 23, 6
- Marston, G. and Watts, R. (2003) Tampering with the evidence: a critical appraisal of evidence-based policy-making, *The Drawing Board: an Australian Review of Public Affairs.*, 3, 3, 143-163
- McKnight, L. and Morgan, A. (2019) The problem with using scientific evidence in education, *EduResearch Matters*, <https://www.aare.edu.au/blog/?p=3874>
- McLaughlin, H., Watts, C. and Beard, M. (2000) Just because it's happening doesn't mean it's working: using action research to improve practice in middle schools, *Phi Delta Kappan*, 82, 4, 284-290
- McLean, R., Graham, D., Tetroe, J. and Volmink, J. (2018) Translating research into action: an international study of the role of research funders, *Health Research Policy and Systems*, 16(44)
- McWilliam, C. (1997) Using a participatory research process to make a difference in policy on aging, *Canadian Public Policy /Analyse de Politiques*, 23, 70-89

- Melnyk, B. (2007) The evidence-based practice mentor: A promising strategy for implementing and sustaining EBP in healthcare systems, *Worldviews on Evidence-Based Nursing*, 4, 3, 123-125
- Menon, A., Korner-Bitensky, N., Kastner, M., McKibbin, K. and Straus, S. (2009) Strategies for rehabilitation professionals to move evidence-based knowledge into practice: a systematic review, *Journal of Rehabilitation Medicine*, 41, 13, 1024-1032
- Miller, D. (1999) The black hole of education research, *Chronicle of Higher Education*, 45, 48, 17-18
- Miller, W., Sorensen, J., Selzer, J. and Brigham, G. (2006) Disseminating evidence-based practices in substance abuse treatment: A review with suggestions, *Journal of Substance Abuse Treatment*, 31, 1, 25-39
- Mosteller, F. and Boruch, R. (2002) *Evidence matters: Randomized trials in education research*, Brookings Institution Press
- Murthy, L., Shepperd, S., Clarke, M., Garner, S., Lavis, J., Perrier, L., Roberts, N. and Straus, S. (2012) Interventions to improve the use of systematic reviews in decision-making by health systems managers, policy makers and clinicians, *Cochrane Databases of Systematic Reviews*, Sept 12 (9).
- Myracle, J. (2019) The hard part about reading, instruction <https://www.edweek.org/ew/articles/2019/05/28/the-hard-part-about-reading-instruction.html?cmp=eml-enl-eu-news1-rm&M=58844468&U=2656422&UUID=2488f7e54f2d4d78ee11c4b8e8337005>
- National Research Council (1999) *Improving student learning: a strategic plan for educational research and its utilization*, Washington DC: National Academy Press
- Nelson, J. and Walker, M. (2019) *Evidence-informed approaches to teaching: Where are we now?*, NFER, <https://www.nfer.ac.uk/news-events/nfer-blogs/evidence-informed-approaches-to-teaching-where-are-we-now/>
- NFER (2017) *The Literacy Octopus*, EEF, <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/the-literacy-octopus-communicating-and-engaging-with-research/>
- Noyes, A. and Adkins, M. (2016) The Impact of Research on Policy: A Case of Qualifications Reform, *British Journal of Educational Studies*, 64:4, 449-465
- Ogden, L., Vinjamuri, M., and Khan, J. (2016) A model for implementing an evidence-based practice in student fieldwork placements, *Journal of Social Service Research*, 42, 4, 425-451
- Oman, K., Duran, C. and Fink, R. (2008) Evidence-based policy and procedures: an algorithm for success, *The Journal of Nursing Administration*, 38, 1, 47-51
- Palmer, J. (1999) Research Matters: A call for the application of empirical evidence to the task of improving the quality and impact of environmental education, *Cambridge Journal of Education*, 29, 3, 379-395
- Pappaioanou, M., Malison, M., Wilkins, K., Otto, B., Goodman, R., Churchill, R., White, M. and Thacker, S. (2003) Strengthening capacity in developing countries for evidence-based public health: the data for decision-making project, *Social Science and Medicine*, 57, 10, 1925-1937
- Parkhurst, J. (2017) *The politics of evidence: From evidence-based policy to the good governance of evidence*, Abingdon, Oxon: Routledge
- Peirson, L., Ciliska, D., Dobbins, M. and Mowat, D. (2012) Building capacity for evidence informed decision making in public health: a case study of organizational change, *BMC Public Health*, 12, 1, 137
- Penuel, W., Briggs, D., Davidson, K., Herlihy, C., Sherer, D., Hill, H., . . . Allen, A. (2017) How school and district leaders access and use research, AERA Open, <https://journals.sagepub.com/doi/abs/10.1177/2332858417705370>
- Perrier L., Mrklas K., Lavis J., Straus S. (2011) Interventions encouraging the use of systematic reviews by health policymakers and managers: a systematic review, *Implementation Science*, 6, 43
- Perry, A., Amadeo, C., Fletcher, M. and Walker, E. (2010) Instinct or Reason: How education policy is made and how we might make it better, CfBT, <https://www.educationdevelopmenttrust.com/~media/EDT/Reports/Research/2010/r-instinct-or-reason-2010.pdf>
- Peterson, M. (2018) In the shadow of politics, *Journal of Health Politics Policy and Law*, 43, 3, 341-376
- Petkovic, J., Welch, V., Jacob, M., Yoganathan, M., Ayala, A., Cunningham, H. and Tugwell, P. (2010) *Do evidence summaries increase health policy-makers' use of evidence from systematic*

- reviews?, Campbell Collaboration, <https://campbellcollaboration.org/library/evidence-summaries-for-policymakers.html>
- Pew Charitable Trusts (2014) *Evidence-based policymaking - A guide for effective government*, <http://www.pewtrusts.org/~media/assets/2014/11/evidencebasedpolicymakingguideforeffectivegovernment.pdf>
- Pogrow, S. (2017) The failure of the US education research establishment to identify effective practices: beware effective practices policies, *Education Policy Analysis Archives*, 25, 5
- Powell, D., Diamond, K., Burchinal, M. and Koehler, M. (2010) Effects of an early literacy professional development intervention on head start teachers and children, *Journal of Educational Psychology*, 102, 2, 299
- Procter, R. (2013) Teachers and research: What they value and what they do, *Journal of Pedagogic Development*, 3, 1
- Public Accounts Committee (2015) *Funding for disadvantaged pupils*, HC 327, 9 October 2015, <https://publications.parliament.uk/pa/cm201516/cmselect/cmpublicacc/327/327.pdf>
- Reid, A. (2016) The use and abuse of research in the public domain, *Australian Educational Researcher*, 43, 1, 75-91
- Roberts, J. (2018) "I might as well be a robot", *TES*, 9/1/18, <https://www.tes.com/news/tes-magazine/tes-magazine/i-might-well-be-a-robot>
- Rohwer, A., Motaze, N., Rehfuess, E. and Young, T. (2017) *E-learning of evidence-based health care (EBHC) to increase EBHC competencies in healthcare professionals: a systematic review*, *Campbell Systematic Reviews*, 4, doi:10.4073/csr.2017.4
- Rolle, I., Zaidi, I., Scharff, J., Jones, D., Firew, A., Enquesselassie, F., Negash, A., Deyessa, N., Mitile, G., Sunderland, N. and Nsubuga, P. (2011) Leadership in strategic information (LSI) building skilled public health capacity in Ethiopia, *BMC Research Notes*, 4, 1, 292
- Rose, J., Thomas, S., Zhang, L., Edwards, A., Augero, A. and Roney, P. (2017) *Research Learning Communities*, [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Research\\_Learning\\_Communities.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Research_Learning_Communities.pdf)
- Rosenbaum, S., Glenton, C. and Oxman, A. (2010) Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information, *Journal of Clinical Epidemiology*, 63, 6, 620-626
- Royal Society/British Academy (2018) *Harnessing educational research*, <https://royalsociety.org/~media/policy/projects/rs-ba-educational-research/educational-research-report.pdf>
- Ruangkanchanasetr, S. (1993) Laboratory investigation utilization in pediatric out-patient department Ramathibodi Hospital, *J Med Assoc Thai*, 76, Suppl 2, 194-208
- Rutter, J. (2012) *Evidence and evaluation in policy-making: A problem of supply or demand?*, Institute for Government
- Schildkamp, K. (2019) Data-based decision-making for school improvement: Research insights and gaps, *Educational Research*, DOI: 10.1080/00131881.2019.1625716
- See, BH, Gorard, S. and Siddiqui, N. (2016) Can teachers use research evidence in practice?: A pilot study of the use of feedback to enhance learning, *Educational Research*, 58, 1, 56-72
- Siddiqi, K., Newell, J. and Robinson, M. (2005) Getting evidence into practice: what works in developing countries?, *International Journal for Quality in Health Care*, 17, 5, 447-454
- Siddiqui, N., Gorard, S. and See, BH (2015) Accelerated Reader as a literacy catch-up intervention during the primary to secondary school transition phase, *Educational Review*, 68, 2, 139-154
- Sirat, M. and Azman, N. (2014) Malaysia's National Higher Education Research Institute (IPPTN): narrowing the research-policy gap in a dynamic higher education system, *Studies in Higher Education*, 39, 8, 1451-1462
- Sisk (1993) Improving the use of research-based evidence in policy making: Effective care in pregnancy and childbirth in the United States, *The Milbank Quarterly*, 71, 3, 477-496
- Slavin, R. (2017) *Education policy in the age of proven school and classroom approaches*, [https://www.huffingtonpost.com/robert-e-slavin/education-policy-in-the-a\\_b\\_13513714.html](https://www.huffingtonpost.com/robert-e-slavin/education-policy-in-the-a_b_13513714.html)
- Slavin, R. and Smith, D. (2009) The relationship between sample sizes and effect sizes in systematic reviews in education, *Educational Evaluation and Policy Analysis*, 31, 4, 500-506

- Slavin, R., Cheung, A., Homes, G., Madden, N. and Chamberlain A. (2013) Effects of a Data-Driven District Reform Model on State Assessments, *AERJ*, 50, 2
- Smith, A. (2003) Scientifically based research and evidence-based education: A Federal Policy context, *Research and Practice for Persons with Severe Disabilities*, 28, 3, 126-132
- Smith, W. (2000) Evidence for the effectiveness of techniques to change physician behaviour, *Chest*, 118, 2, 8S-17S
- Sparks, S. (2018) Teachers want education research, *Education Week*, 27<sup>th</sup> November, [http://blogs.edweek.org/edweek/inside-school-research/2018/11/get\\_teachers\\_research\\_on\\_education.html?cmp=RSS-FEED](http://blogs.edweek.org/edweek/inside-school-research/2018/11/get_teachers_research_on_education.html?cmp=RSS-FEED)
- Speight, S., Callahan, M., Griggs, J. and Javiera, C. (2016) *Rochdale research into practice: evaluation report and executive summary*, London: Education Endowment Foundation
- Stang, A., Poole, C. and Kuss, O. (2010) The ongoing tyranny of statistical significance testing in biomedical research, *European Journal of Epidemiology*, 25, 4, 225-230
- Starbuck, W. (2016) 60th Anniversary Essay: How journals could improve research practices in social science, *Administrative Science Quarterly*, 61, 2, 165–183
- Stecher, B., Holtzman, D., Garet, M., Hamilton, L., Engberg, J., Steiner, E., Robyn, A., Baird, M., Gutierrez, I., Peet, E., de los Reyes, I., Fronberg, K., Weinberger, G. Hunter, G. and Chambers, J. (2019) *Improving Teaching Effectiveness: Final Report*, Rand, [https://www.rand.org/pubs/research\\_reports/RR2242.html](https://www.rand.org/pubs/research_reports/RR2242.html)
- Strambler, M. and McKown, C. (2013) Promoting student engagement through evidence-based action research with teachers, *Journal of Educational and Psychological Consultation*, 23, 2, 87-114
- Taylor, C. and Gorard, S. (2002) *The RBCN Consultation exercise: Stakeholder Report*, Occasional Paper 50, Cardiff University School of Social Sciences
- Taylor, R., Reeves, B., Ewings, P. and Taylor, R. (2004) Critical appraisal skills training for health care professionals: a randomized controlled trial, *BMC Medical Education*, 4, 30–40
- The Guardian (2017) David Laws: “the quality of education policy-making is poor”, *The Guardian*, 1/8/17, <https://www.theguardian.com/education/2017/aug/01/david-laws-education-policy-schools-minister-thinktank-epi>
- Thomas I., Mackie R., Sheldrick, C., Hyde, J. and Leslie, L. (2015) Exploring the Integration of Systems and Social Sciences to Study Evidence Use among Child Welfare Policy-makers, *Child Welfare*, 94, 3, 33-58
- Thompson D., Estabrooks C., Scott-Findlay S., Moore K. and Wallin L. (2007) Interventions aimed at increasing research use in nursing: a systematic review, *Implementation Science*, 2, 15
- Thomson, P., Angus, N. and Scott, J. (2000) Building a framework for getting evidence into critical care education and practice, *Intensive and Critical Care Nursing*, 16, 3, 164-174
- Tranmer J., Lochaus-Gerlach J., Lam M. (2002) The effect of staff nurse participation in a clinical nursing research project on attitude towards, access to, support of and use of research in acute care settings, *Canadian Journal of Nurse Leadership*, 15, 18-26
- Traynor, R., DeCorby, K. and Dobbins, M. (2014) Knowledge brokering in public health: a tale of two studies, *Public Health*, 128, 6, 533-544
- Uneke, C., Ezeoha, A., Ndukwe, C., Oyibo, P. and Onwe, F. (2012) Promotion of evidence-informed health policymaking in Nigeria: bridging the gap between researchers and policymakers, *Global Public Health*, 7, 7, 750-765
- Uneke, C., Ezeoha, A., Uro-Chukwu, H., Ezeonu, C. and Igboji, J. (2018) Promoting researchers and policy-makers collaboration in evidence-informed policy-making in Nigeria, *International Journal of Health Policy Management*, 7, 6, 522-531
- Uneke, C., Ezeoha, A., Uro-Chukwu, H., Ezeonu, C., Ogbu, O., Onwe, F., Edoga, C. (2015a) Enhancing the capacity of policy-makers to develop evidence-informed policy brief on infectious diseases of poverty in Nigeria, *International Journal of Health Policy and Management*, 4, 9, 599-610
- Uneke, C., Ndukwe, C., Ezeoha, A., Uro-Chukwu, H., Ezeonu, C. (2015b) Implementation of a health policy advisory committee as a knowledge translation platform: the Nigeria experience, *International Journal of Health Policy Management*, 4, 3, 161–168
- University of Bristol (2017) Research Learning Communities, <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/research-learning-communities/>



- Urwick, J. (2014) Who will listen to educational researchers? Reflections on the research–policy relationship in the global South, with illustrations from Africa, *Compare*, 44, 4, 545-565
- Valentine, A., DeAngelo, D., Alegría, M. and Cook, B. (2014) Translating disparities research to policy: A qualitative study of state mental health policymakers' perceptions of mental health care disparities report cards, *Psychological Services*, 11, 4, 377-387
- van de Arend, J. (2016) *Connecting research and policy: how linkages support the use of research evidence in social policymaking*, (PhD), The University of Queensland, [https://espace.library.uq.edu.au/data/UQ\\_411282/s42526993\\_final\\_thesis.pdf?Expires=1531784801&Signature=d3zjHuBD~1W2hSz6R2jynWTHAsKWXjNTykQjq0Qaj3zzY2zhF991xLSx6naRAEGQNameW50q~-tWb-pmZQVIY2JR6hAQH9LCLT-PlbOq5R8dUO6HUQWznRkoeqrtgALhV9pFxFKArgpVh9BCWoFhPQEyP5aUaOcb6gIm5qTUBFOgrwXeBfPumlVF4Y7OgH4Ig8HnEN~KUCmPTUCpurAmpjkoA5clJJ~IVqK2VFv-tM-Gm4lt76JUNcWGAkEHCPiq9UzJs-eyWeKCxWvnw09XlcIp5XFtJKLZlv6NpmVsaZBsPorxHIHt61mXXnUkOJHNT7i7s9KG6eCzqPILQZ07Ug\\_\\_&Key-Pair-Id=APKAJKNB4MJBNC6NLQ](https://espace.library.uq.edu.au/data/UQ_411282/s42526993_final_thesis.pdf?Expires=1531784801&Signature=d3zjHuBD~1W2hSz6R2jynWTHAsKWXjNTykQjq0Qaj3zzY2zhF991xLSx6naRAEGQNameW50q~-tWb-pmZQVIY2JR6hAQH9LCLT-PlbOq5R8dUO6HUQWznRkoeqrtgALhV9pFxFKArgpVh9BCWoFhPQEyP5aUaOcb6gIm5qTUBFOgrwXeBfPumlVF4Y7OgH4Ig8HnEN~KUCmPTUCpurAmpjkoA5clJJ~IVqK2VFv-tM-Gm4lt76JUNcWGAkEHCPiq9UzJs-eyWeKCxWvnw09XlcIp5XFtJKLZlv6NpmVsaZBsPorxHIHt61mXXnUkOJHNT7i7s9KG6eCzqPILQZ07Ug__&Key-Pair-Id=APKAJKNB4MJBNC6NLQ)
- Van de Goor, I., Hamalainen, R., Syed, A., Lau, C., Sandu, P., Spitters, H., . . . and Aro, A. R. (2017) Determinants of evidence use in public health policymaking: results from a study across six EU countries, *Health Policy*, 121, 3, 273-281
- Vujcich, D. (2016) Where there is no evidence, and where evidence is not enough: an analysis of policy-making to reduce the prevalence of Australian indigenous smoking, *Frontiers in Public Health*, 4, 228
- Walker, M., Nelson, J., Bradshaw, S. with Brown, C (2019) Teachers' engagements with research; what do we know? A research briefing, London: Education Endowment Foundation
- Waq, G., Mavo, H., Snowdon, W., Moodie, M., Schultz, J., McCabe, M., Kremer, P. and Swinburn, B. (2013) Knowledge brokering between researchers and policymakers in Fiji to develop policies to reduce obesity: a process evaluation, *Implementation Science*, 8, 1, 74
- Ward, V., House, A. and Hamer, S. (2009a) Knowledge brokering: exploring the process of transferring knowledge into action, *BMC Health Services Research*, 9, 1, 12
- Ward, V., House, A. and Hamer, S. (2009b) Knowledge brokering: the missing link in the evidence to action chain?, *Evidence and Policy*, 5, 3, 267-279
- Wentworth, L., Mazzeo, C. and Connolly, F. (2017) Research practice partnerships: A strategy for promoting evidence-based decision-making in education, *Educational Research*, 59, 2, 241-255
- Wexler, N. (2019) *Three reasons classroom practice conflicts with evidence*, Forbes, <https://www.forbes.com/sites/nataliewexler/2019/08/03/three-reasons-classroom-practice-conflicts-with-evidence-on-how-kids-learn/#71114cad7b0a>
- Wiggins, M., Jerrim, J., Tripney, J., Khatwa, M. and Gough, D. (2019) The RISE project: evidence-informed
- Williams, A. (2010) Is Evidence-Based Policy Making Really Possible?, in Colebatch, H., Hoppe, R. and Noordegraaf, M. (Eds.) *Working for Policy* (pp. 195-210), Amsterdam: Amsterdam University Press
- Williams, M. (2016) *To what extent has research been used to inform anti-poverty policy in Ghana*, (Electronic Thesis or Dissertation), <http://search.ebscohost.com/login.aspx?direct=true&db=ddu&AN=A2CB93E394CC211B&site=ehost-live> Available from EBSCOhost ddu database
- Winters, C., and Echeverri, R. (2012) Teaching strategies to support evidence-based practice, *Critical Care Nurse*, 32, 3, 49-54
- Winters, C., and Echeverri, R. (2012) Teaching strategies to support evidence-based practice, *Critical Care Nurse*, 32, 3, 49-54
- Wolgemuth, J., Hicks, T. and Agosto, V. (2017) Unpacking assumptions in research synthesis, *Educational Researcher*, 46, 3, 131-139
- Wyatt, J., Paterson-Brown, S., Johanson, R., Altman, D., Bradburn, M. and Fisk, N. (1998) Randomised Trial of educational visits to enhance use of systematic reviews in 25 obstetric units, *BMJ*, 317, 7165, 1041-1046

- Xiao, Z., Kasim, A. and Higgins, S. (2016) Same difference? Understanding variation in the estimation of effect sizes from trial data, *International Journal of Educational Research*, 77, 1-14, <http://www.sciencedirect.com/science/article/pii/S0883035515304808>
- Yost J., Ganann R., Thompson D., Aloweni F., Newman K., Hazzan A., McKibbin A., Dobbins M. and Ciliska D. (2015) The effectiveness of knowledge translation interventions for promoting evidence-informed decision-making among nurses in tertiary care: a systematic review and meta-analysis, *Implementation Science*, 10, 1, 1-15
- Zimmermann, K. (2004) Advising policymakers through the media, *The Journal of Economic Education*, 35, 4, 395-406
- Zlotnik, J. (2007) Evidence-based practice and social work education: A view from Washington, *Research on Social Work Practice*, 17, 5, 625-629

---

The operation of the sieve is explained fully in the reference cited in the text. Users read each column in turn from the left, and move down (never up) the rows until the research under consideration is at least as good as the description for that column is. They then move to the next column, remaining in the same row, and repeat the process. Using this simple approach, in discussions at the outset and during other projects, the three raters never disagreed on the rating for a piece.

The sieve has been deliberately created without arbitrary thresholds. It is possible to create such thresholds, perhaps counting 400 cases or more per cell as being large and 20 cases or less as being very small, for example. Or counting below 2% attrition as being low, and above 50% as being very high. But these exact figures have a false precision. Not only are they arbitrary, they take no account of key analytical factors such as the balance, type and likely cause of attrition in each group. A loss of 3% data from each group in a comparison might be a small amount if it were both balanced between groups and had the same apparent cause. A loss of 3% might be problematic if it were created by resentment among the control that they did not receive the treatment, and resentment among the treatment group that what they may be required to do is patronising. This kind of differential reason for dropout is something we have met in practice (Gorard et al. 2016).

The allocation of a quality grade to any piece of research is, of course, a judgement. But allocating fixed threshold numbers to the sieve would also involve judgement. The sieve encourages the judgement to be made on the basis of evidence, rather than *a priori* and independently of other contextual factors (as in the examples above).

Some commentators might believe that scale, for example, can be pre-stated more precisely using power calculations. But such calculations of scale are circular, based on assuming a specific 'effect' size being sought in a future evaluation (and so, indirectly, assuming the scale). Also power calculations are defined in terms of the conduct of significance tests which are outdated, misleading, and hardly ever relevant because real-life data is usually incompletely randomised. In fact, using the notion of power alongside a measure of attrition (where any attrition creates non-random allocation by definition) is a mathematical contradiction.

For the purposes of this paper, we sought research that could address our causal question. This means that we agreed only to include designs that had some kind of counterfactual, even if only a convenient comparison group. A true RCT would be rated as 4 🗝️, and a convenient comparison group would be rated as 1 🗝️, with a quasi-experimental design at 2 🗝️ or 3 🗝️, as explained in Gorard et al. (2017), along with all of the other columns in the sieve. As another example, where the outcome measure was a simple count or standardised test score or similar, the research could remain at whatever grade it was, but if the outcome measure was a latent/attitude variable or a non-standardised score then a 3 🗝️ was the maximum grade permitted.