

Evidence-Based Educational Practice

Summary

Evidence-based practice (EBP) is a buzzword in contemporary professional debates, for example, in education, medicine, psychiatry, and social policy. It is known as the “what works” agenda, and its focus is on the use of the best available evidence to bring about desirable results or prevent undesirable ones. We immediately see here that EBP is practical in nature, that evidence is thought to play a central role, and also that EBP is deeply causal: we intervene into an already existing practice in order to produce an output or to improve the output. If our intervention brings the results we want, we say that it “works.”

How should we understand the causal nature of EBP be understood? Causality is a highly contentious issue in education, and many writers want to banish it altogether. But causation denotes a dynamic relation between factors and is indispensable if one wants to be able to plan the attainment of goals and results. A nuanced and reasonable understanding of causality is therefore necessary to EBP, and this we find in the INUS-condition approach.

The nature and function of evidence is much discussed. The evidence in question is supplied by research, as a response to both political and practical demands that educational research should contribute to practice. In general, evidence speaks to the truth value of claims. In the case of EBP, the evidence emanates from randomized controlled trials (RCTs) and presumably speaks to the truth value of claims such as “if we do X, it will lead to result Y.” But what does research evidence really tell us? It is argued here that a positive RCT result will tell you that X worked where the RCT was conducted and that an RCT does not yield general results.

Causality and evidence come together in the practitioner perspective. Here we shift from *finding* causes to *using* them to bring about desirable results. This puts contextual matters at center stage: will X work in this particular context? It is argued that much heterogeneous contextual evidence is required to make X relevant for new contexts. If EBP is to be a success, research evidence and contextual evidence must be brought together.

Keywords

evidence, causality, context, effectiveness, INUS conditions, RCT, relevance, practitioner perspective, what works

Introduction

Evidence-based practice, hereafter EBP, is generally known as the “what works” agenda. This is an apt phrase, pointing as it does to central practical issues: how to attain goals and produce desirable results, and how we know what works. Obviously, this goes to the heart of much (but not all) of the everyday activity that practitioners engage in. The “what works” agenda is meant to narrow the gap between research and practice and be an area in which research can make itself directly useful to practice. David Hargreaves, one of the instigators of the EBP debate in education, has stated that the point of evidence-based research is to gather evidence about what works in what circumstances (Hargreaves, 1996a, 1996b). Teachers, Hargreaves said, want to know what works; only secondarily are they interested in understanding the why of classroom events. The kind of research we talk about is meant to be relevant not only for teachers but also for policymakers, school developers, and headmasters. Its purpose is to improve practice, which largely comes down to improving student achievement. Hargreaves’s work was supported by, for

example, Robert Slavin, who stated that education research not only can address questions about “what works” but also must do so (Slavin, 2004).

All the same, despite the fact that EBP, at least at the outset, seems to speak directly to the needs of practitioners, it has met with much criticism. It is difficult to characterize both EBP and the debate about it, but let me suggest that the debate branches off in different but interrelated directions. We may roughly identify two: what educational research can and should contribute to practice and what EBP entails for the nature of educational practice and the teaching profession. There is ample space here for different definitions, different perspectives, different opinions, as well as for some general unclarity and confusions. To some extent, advocates and critics bring different vocabularies to the debate, and to some extent, they employ the same vocabulary but take very different stances. Overall in the EBP conceptual landscape we find such concepts as relevance, effectiveness, generality, causality, systematic reviews, randomized controlled trials (RCTs), what works, accountability, competences, outcomes, measurement, practical judgment, professional experience, situatedness, democracy, appropriateness, ends, and means as constitutive of ends or as instrumental to the achievement of ends. Out of this tangle we shall carefully extract and examine a selection of themes, assumptions, and problems. These mainly concern the causal nature of EBP, the function of evidence, and EBP from the practitioner point of view.

Definition, History, and Context

The term “evidence-based” originates in medicine—evidence-based medicine—and was coined in 1991 by a group of doctors at McMaster University in Hamilton, Ontario. Originally, it denoted a method for teaching medicine at the bedside. It has long since outgrown the hospital

bedside and has become a buzzword in many contemporary professions and professional debates, not only in education, but also leadership, psychiatry, and policymaking. The term EBP can be defined in different ways, broadly or more narrowly. We shall here adopt a parsimonious, minimal definition, which says that *EBP involves the use of the best available evidence to bring about desirable outcomes, or conversely, to prevent undesirable outcomes* (Kvernbekk, 2016).

That is to say, we intervene to bring about results, and this practice should be guided by evidence of how well it works. This minimal definition does not specify what kinds of evidence are allowed, what “based” should mean, what practice is, or how we should understand the causality that is inevitably involved in bringing about and preventing results. Minimal definitions are eminently useful because they are broad in their phenomenal range and thus allow differing versions of the phenomenon in question to fall under the concept.

We live in an age which insists that practices and policies of all kinds be based on research. Researchers thus face political demands for better research bases to underpin, inform and guide policy and practice, and practitioners face political demands to make use of research to produce desirable results or improve results already produced. Although the term EBP is fairly recent, the idea that research should be used to guide and improve practice is by no means new. To illustrate, in 1933, the *School Commission of Norwegian Teacher Unions* (Lærerorganisasjonenes skolenavn, 1933) declared that progress in schooling can only happen through empirical studies, notably, by different kinds of experiments and trials. Examples of problems the commission thought research should solve are (a) in which grade the teaching of a second language should start and (b) what the best form of differentiation is. The accumulated evidence should form the basis for policy, the unions argued. Thus, the idea that pedagogy should be based on systematic research is not entirely new. What is new is the magnitude and

influence of the EBP movement and other, related trends, such as large-scale international comparative studies (e.g., the Progress in International Reading Literacy Study, PIRLS, and the Programme for International Student Assessment, PISA). Schooling is generally considered successful when the predetermined outcomes have been achieved, and education worldwide therefore makes excessive requirements of assessment, measurement, testing, and documentation. EBP generally belongs in this big picture, with its emphasis on knowing what works in order to maximize the probability of attaining the goal. What is also new, and quite unprecedented, is the growth of organizations such as the What Works Clearinghouses, set up all around the world. The WWCs collect, review, synthesize, and report on studies of educational interventions. Their main functions are, first, to provide hierarchies that rank evidence. The hierarchies may differ in their details, but they all rank RCTs, meta-analyses, and systematic reviews on top and professional judgment near the bottom (see, e.g., [Oancea & Pring, 2008](#)). Second, they provide guides that offer advice about how to choose a method of instruction that is backed by good evidence; and third, they serve as a warehouse, where a practitioner might find methods that are indeed backed by good evidence ([Cartwright & Hardie, 2012](#)).

Educationists today seem to have a somewhat ambiguous relationship to research and what it can do for practice. Some, such as Robert [Slavin \(2002\)](#), a highly influential educational researcher and a defender of EBP, think that education is on the brink of a scientific revolution. Slavin has argued that over time, rigorous research will yield the same step-by-step, irreversible progress in education that medicine has enjoyed because all interventions would be subjected to strict standards of evaluation before being recommended for general use. Central to this optimism is the RCT. Other educationists, such as Gert [Biesta \(2007, 2010\)](#), also a highly influential figure in the field and a critic of EBP, are wary of according such weight to research

and to the advice guides and practical guidelines of the WWCs for fear that this might seriously restrict, or out and out replace, the experience and professional judgment of practitioners. And there matters stand: EBP is a huge domain with many different topics, issues, and problems, where advocates and critics have criss-crossing perspectives, assumptions, and value stances.

The Causal Nature of Evidence-Based Practice

As the slogan “what works” suggests, EBP is practical in nature. By the same token, EBP is also deeply causal. *Works* is a causal term, as are *intervention*, *effectiveness*, *bring about*, *influence*, and *prevent*. In EBP we intervene into an already existing practice in order to change its outcomes in what we judge to be a more desirable direction. To say that something (an intervention) works is roughly to say that doing it yields the outcomes we want. If we get other results or no results at all, we say that it does not work. To put it crudely, we do X, and if it leads to some desirable outcome Y, we judge that X works. It is the ambition of EBP to provide knowledge of how intervention X can be used to bring about or produce Y (or improvements in Y) and to back this up by solid evidence—for example, how implementing a reading-instruction program can improve the reading skills of slow or delayed readers, or how a schoolwide behavioral support program can serve to enhance students’ social skills and prevent future problem behavior. For convenience, I adopt the convention of calling the cause (intervention, input) X and the effect (result, outcome, output) Y. This is on the explicit understanding that both X and Y can be highly complex in their own right, and that the convention, as will become clear, is a simplification.

There can be no doubt that EBP is causal. However, the whole issue of causality is highly contentious in education. Many educationists and philosophers of education have over the years

dismissed the idea that education is or can be causal or have causal elements. In EBP, too, this controversy runs deep. By and large, advocates of EBP seem to take for granted that causality in the social and human realm simply exists, but they tend not to provide any analysis of it. RCTs are preferred because they allow causal inferences to be made with a high degree of certainty. As Slavin (2002) put it, “The experiment is the design of choice for studies that seek to make causal conclusions, and particularly for evaluations of educational innovations” (p. 18). In contrast, critics often make much of the causal nature of EBP, since for many of them this is reason to reject EBP altogether. Biesta is a case in point. For him and many others, education is a moral and social practice and therefore *noncausal*. According to Biesta (2010):

The most important argument against the idea that education is a causal process lies in the fact that education is not a process of physical interaction but a process of *symbolic* or *symbolically mediated* interaction. (p. 34)

Since education is noncausal and EBP is causal, on this line of reasoning, it follows that EBP must be rejected—it fundamentally mistakes the nature of education.

Such wholesale dismissals rest on certain assumptions about the nature of causality, for example, that it is deterministic, positivist, and physical and that it essentially belongs in the natural sciences. Biesta, for example, clearly assumes that causality requires a physical process. But since the mid-1900s our understanding of causality has witnessed dramatic developments; arguably the most important of which is its reformulation in probabilistic terms, thus making it compatible with indeterminism. A quick survey of the field reveals that causality is a highly varied thing. The concept is used in different ways in different contexts, and not all uses are compatible. There are several competing theories, all with counterexamples. As Nancy Cartwright (2007b) has pointed out, “There is no single interesting characterizing feature of

causation; hence no off-the-shelf or one-size-fits-all method for finding out about it, no ‘gold standard’ for judging causal relations” (p. 2).

The approach to causality taken here is twofold. First, there should be room for causality in education; we just have to be very careful how we think about it. Causality is an important ingredient in education because it denotes a dynamic relationship between factors of various kinds. Causes make their effects happen; they make a difference to the effect. Causality implies change and how it can be brought about, and this is something that surely lies at the heart of education. Ordinary educational talk is replete with causal verbs, for example, enhance, improve, reduce, increase, encourage, motivate, influence, affect, intervene, bring about, prevent, enable, contribute. The short version of the causal nature of education, and so EBP, is therefore that EBP is causal because it concerns the bringing about of desirable results (or the preventing of undesirable results). We have a causal connection between an action or an intervention and its effect, between X and Y. The longer version of the causal nature of EBP takes into account the many forms of causality: direct, indirect, necessary, sufficient, probable, deterministic, general, actual, potential, singular, strong, weak, robust, fragile, chains, multiple causes, two-way connections, side-effects, and so on. What is important is that we adopt an understanding of causality that fits the nature of EBP and does not do violence to the matter at hand. That leads me to my second point: the suggestion that in EBP causes are best understood as INUS conditions.

The understanding of causes as INUS conditions was pioneered by the philosopher John Mackie (1975). He placed his account within what is known as the regularity theory of causality. Regularity theory is largely the legacy of David Hume, and it describes causality as the constant conjunction of two entities (cause and effect, input and output). Like many others, Mackie took (some version of) regularity theory to be the common view of causality. Regularities are

generally expressed in terms of necessity and sufficiency. In a causal law, the cause would be held to be both necessary and sufficient for the occurrence of the effect; the cause would produce its effect every time; and the relation would be constant. This is the starting point of Mackie's brilliant refinement of the regularity view. Suppose, he said, that a fire has broken out in a house, and that the experts conclude that it was caused by an electrical short circuit. How should we understand this claim? The short circuit is not necessary, since many other events could have caused the fire. Nor is it sufficient, since short circuits may happen without causing a fire. But if the short circuit is neither necessary nor sufficient, then what do we mean by saying that it caused the fire? What we mean, Mackie (1975) suggests, is that the short circuit is an INUS condition: "an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result" (p. 16), INUS being an acronym formed of the initial letters of the italicized words. The main point is that a short circuit does not cause a fire all by itself; it requires the *presence* of oxygen and combustible material and the *absence* of a working sprinkler. On this approach, therefore, a cause is a complex set of conditions, of which some may be positive (present), and some may be negative (absent). In this constellation of factors, the event that is the focus of the definition (the insufficient but necessary factor) is the one that is salient to us. When we speak of an event causing another, we tend to let this factor represent the whole complex constellation.

In EBP, our own intervention X (strategy, method of instruction) is the factor we focus on, the factor that is salient to us, is within our control, and receives our attention. I propose that we understand any intervention we implement as an INUS condition. Then it immediately transpires that X not only *does not* bring about Y alone, but also that it *cannot* do so.

Before inquiring further into interventions as INUS -conditions, we should briefly characterize causality in education more broadly. Most causal theories, but not all of them, understand causal connections in terms of probability—that is, causing is making more likely. This means that causes sometimes make their effects happen, and sometimes not. A basic understanding of causality as indeterministic is vitally important in education, for two reasons. First, because the world is diverse, it is to some extent unpredictable, and planning for results is by no means straightforward. Second, because we can here clear up a fundamental misunderstanding about causality in education: causality is not deterministic and the effect is therefore not necessitated by the cause. The most common phrase in causal theory seems to be that causes make a difference for the effect (Schaffer, 2007). We must be flexible in our thinking here. One factor can make a difference for another factor in a great variety of ways: prevent it, contribute to it, enhance it as part of a causal chain, hinder it via one path and increase it via another, delay it, or produce undesirable side effects, and so on. This is not just conceptual hair-splitting; it has great practical import. Educational researchers may tell us that X causes Y, but what a practitioner can do with that knowledge differs radically if X is a potential cause, a disabler, a sufficient cause, or the absence of a hindrance.

Interventions as INUS Conditions

Human affairs, including education, are complex, and it stands to reason that a given outcome will have several sources and causes. While one of the factors in a causal constellation is salient to us, the others jointly enable X to have an effect. This enabling role is eminently generalizable and crucial to understanding how interventions bring about their effects. As Mackie's example

suggests, enablers may also be absences—that is vital to note, since absences normally go under our radar.

The term “intervention” deserves brief mention. To some it seems to denote a form of practice that is interested only (or mainly) in producing measurable changes on selected output variables. It is not obvious that there is a clear conception of intervention in EBP, but we should refrain from imposing heavy restrictions on it. I thus propose to employ the broad understanding suggested by [Peter Menzies and Huw Price \(1993\)](#)—namely, interventions as a natural part of human agency. We all have the ability to intervene in the world and influence it; that is, to act as agents. Educational interventions may thus take many forms and encompass actions, strategies, programs and methods of instruction. Most interventions will be composites consisting of many different activities, and some, for instance, schoolwide behavioral programs, are meant to run for a considerable length of time.

When practitioners consider implementing an intervention X, the INUS approach encourages them to also consider what the enabling conditions are and how they might allow X to produce Y (or to contribute to its production). Our general knowledge of house fires and how they start prompts us to look at factors such as oxygen, materials, and fire extinguishers. In other cases, we might not know what the enabling conditions are. Suppose a teacher observes that some of his first graders are reading delayed. What to do? The teacher may decide to implement what we might call “Hatcher’s method” ([Hatcher et al., 2006](#)). This “method” focuses on letter knowledge, single-word reading, and phoneme awareness and lasts for two consecutive 10-week periods. Hatcher and colleagues’ study showed that about 75% of the children who received it made significant progress. So should our teacher now simply implement the method and expect the results with his own students to be (approximately) the same? As any teacher knows, what

worked in one context might not work in another context. What we can infer from the fact that the method, X, worked where the data were collected is that a sufficient set of support factors were present to enable X to work. That is, Hatcher's method serves as an INUS condition in a larger constellation of factors that together are sufficient for a positive result for a good many of the individuals in the study population. Do we know what the enabling factors are—the factors that correspond to presence of oxygen and inflammable material and absence of sprinkler in Mackie's example? Not necessarily. General educational knowledge may tell us something, but enablers are also contextual. Examples of possible enablers include student motivation, parental support (important if the method requires homework), adequate materials, a separate room, and sufficient time. Maybe the program requires a teacher's assistant? The enablers are factors that X *requires* to bring about or improve Y; if they are missing, X might not be able to do its work.

Understanding X as an INUS condition adds quite a lot of complexity to the simple X–Y picture and may thus alleviate at least some of the EBP critics' fear that EBP is inherently reductionist and oversimplified. EBP is at heart causal, but that does not entail a deterministic, simplistic or physical understanding. Rather, I have argued, to do justice to EBP in education its causal nature must be understood to be both complex and sophisticated. We should also note here that X can enter into different constellations. The enablers in one context need not be the same as the enablers in another context. In fact, we should expect them to be different, simply because contexts are different.

Evidence and Its Uses

Evidence is an epistemological concept. In its immediate surroundings we find such concepts as justification, support, hypotheses, reasons, grounds, truth, confirmation, disconfirmation,

falsification, and others. It is often unclear what people take evidence and its function to be. In epistemology, evidence is that which serves to confirm or disconfirm a hypothesis (claim, belief, theory; [Achinstein, 2001](#); [Kelly, 2008](#)). The basic function of evidence is thus summed up in the word “support”: evidence is something that stands in a relation of support (confirmation, disconfirmation) to a claim or hypothesis, and provides us with good reason to believe that a claim is true (or false). The question of what *can count as* evidence is the question of what kind of stuff can enter into such evidential relations with a claim. This question is controversial in EBP and usually amounts to criticism of evidence hierarchies. The standard criticisms are that such hierarchies unduly privilege certain forms of knowledge and research design ([Oancea & Pring, 2008](#)), undervalue the contribution of other research perspectives ([Pawson, 2012](#)), and undervalue professional experience and judgment ([Hammersley, 1997, 2004](#)). It is, however, not of much use to discuss evidence in and of itself—we must look at what we want evidence *for*. Evidence is that which can perform a support function, including all sorts of data, facts, personal experiences, and even physical traces and objects. In murder mysteries, bloody footprints, knives, and witness observations count as evidence, for or against the hypothesis that the butler did it. In everyday life, a face covered in ice cream is evidence of who ate the dessert before dinner.

There are three important things to keep in mind concerning evidence. First, in principle, many different entities can play the role of evidence and enter into an evidentiary relation with a claim (hypothesis, belief). Second, what counts as evidence in each case has everything to do with the type of claim we are interested in. If we want evidence that something is possible, observation of one single instance is sufficient evidence. If we want evidence for a general claim, we at least need enough data to judge that the hypothesis has good inductive support. If we want

to bolster the normative conclusion that means M1 serves end E better than means M2, we have to adduce a range of evidences and reasons, from causal connections to ethical considerations (Hitchcock, 2011). If we want to back up our hypothesis that the butler is guilty of stealing Lady Markham's necklace, we have to take into consideration such diverse pieces of evidence as fingerprints, reconstructed timelines, witness observations and alibis. Third, evidence comes in different degrees of trustworthiness, which is why evidence must be evaluated—bad evidence cannot be used to support a hypothesis and does not speak to its truth value; weak evidence can support a hypothesis and speak to its truth value, but only weakly.

The goal in EBP is to find evidence for a causal claim. Here we meet with a problem, because causal claims come in many different shapes: for example, “X leads to Y,” “doing X sometimes leads to Y and sometimes to G,” “X contributes moderately to Y” and “given Z, X will make a difference to Y.” On the *INUS* approach the hypothesis is that X, in conjunction with a suitable set of support factors, in all likelihood will lead to Y (or will contribute positively to Y, or make a difference to the bringing about of Y). The reason why RCTs are preferred is precisely that we are dealing with causal claims. Provided that the RCT design satisfies all requirements, it controls for confounders, and makes it possible to distinguish correlations from causal connections and to draw causal inferences with a high degree of confidence. In RCTs we compare two groups, the study group and the control group. Random assignment is supposed to ensure that the groups have the same distribution of causal and other factors, save one—namely, the intervention X (but do note that the value of randomization has recently been problematized, most notably by John Worrall (2007)). The standard result from an RCT is a treatment effect, expressed in terms of an effect size. An effect size is a statistical measure denoting average effect in the treatment group minus average effect in the control group (to simplify). We tend to assume

that any difference between the groups requires a causal explanation. Since other factors and confounders are (assumed to be) evenly distributed and thus controlled for, we infer that the treatment, whatever it is, is the cause of the difference. Thus, the evidence-ranking schemes seem to have some justification, despite Cartwright's insistence that there is no gold standard for drawing causal inferences. We want evidence for causal claims, and RCTs yield highly trustworthy evidence and, hence, give us good reason to believe the causal hypothesis. In most cases the causal hypothesis is of the form "if we do X it will lead to Y."

Effectiveness

Effectiveness is much sought after in EBP. For example, Philip Davies (2004) describes the role of the Campbell Collaboration as helping both policymakers and practitioners make good decisions by providing systematic reviews of the effectiveness of social and behavioral interventions in education. The US Department of Education's *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide* (2003) provides an example of how evidence, evidence hierarchies, effectiveness, and "what works" are tied together. The aim of the guide is to provide practitioners with the tools to distinguish practices that are supported by rigorous evidence from practices that are not. "Rigorous evidence" is here identical to RCT evidence, and the guide devotes an entire chapter to RCTs and why they yield strong evidence for the effectiveness of some intervention. Thus:

- The intervention should be demonstrated effective, through well-designed randomized controlled trials, in more than one site of implementation;
- These sites should be typical school or community settings, such as public school classrooms taught by regular teachers; and

- The trials should demonstrate the intervention’s effectiveness in school settings similar to yours, before you can be confident that it will work in your schools/classrooms (p. 17).

Effectiveness is clearly at the heart of EBP, but what does it really mean? “Effectiveness” is a complex multidimensional concept containing causal, normative, and conceptual dimensions, all of which have different sides to them. Probabilistic causality comes in two main versions, one concerning causal *strength* and one concerning causal *frequency* or *tendency* (Kvernbekk, 2016). One common interpretation says that effectiveness concerns the relation between input and output—that is, the *degree* to which an intervention works. Effect sizes would seem to fall into this category, expressing as they do the magnitude of the effect and thereby the *strength* of the cause.

But a large effect size is not the only thing we want; we also want the cause to make its effect happen regularly across different contexts. In other words, we are interested in *frequency*. A cause may not produce its effect every time but often enough to be of interest. If we are to be able to plan for results, X must produce its effect regularly. Reproducibility of desirable results thus depends crucially of the tendency of the cause to produce its effect wherever and whenever it appears. Hence, the term “effectiveness” signals generality. In passing, the same generality hides in the term “works”—if an intervention works, it can be relied on to produce its desired results wherever it is implemented. The issue of scope also belongs to this generality picture: for which groups do we think our causal claim holds? All students of a certain kind, for example, first graders who are responsive to extra word and phoneme training? Some first graders somewhere in the world? All first graders everywhere?

The normative dimension of “what works,” or effectiveness, is equally important, also because it demonstrates so well that effectiveness is a judgment we make. We sometimes gauge

effectiveness by the relation between desired output and actual output; that is, if the correlation between the two is judged to be sufficiently high, we conclude that the method of instruction in question is effective. In such cases, the result (actual or desired) is central to our judgment, even if the focus of EBP undeniably lies on the means, and not on the goals. In a similar vein, to conclude that X works, you must judge the output to be satisfactory (enough), and that again depends on which success criteria you adopt (Morrison, 2001). Next, we have to consider the temporal dimension: how long must an effect linger for us to judge that X works? Three weeks? Two months? One year? Indefinitely? Finally, there is a conceptual dimension to judgments of effectiveness: the judgment of how well X works also depends on how the target is defined. For example, an assessment of the effectiveness of reading-instruction methods depends on what it means to say that students can read. Vague target articulations give much leeway for judgments of whether the target (Y) is attained, which, in turn, opens the possibility that many different Xs are judged to lead to Y.

Given the different dimensions of the term “effectiveness,” we should not wonder that effectiveness claims often equivocate on whether they mean effectiveness in terms of strength or frequency or perhaps both. The intended scope is often unclear, the target may be imprecise and the success criteria too broad or too narrow or left implicit altogether. However, since reproducibility of results is vitally important in EBP, it stands to reason that generality—external validity—should be of the greatest interest. All the strategies Dean, Hubbell, Pitler, and Stone (2012) discuss in their book about classroom instruction that works are explicitly general, for example, that providing feedback on homework assignments will benefit students and help enhance their achievements. This is generality in the frequency and (large) scope sense. It is

future oriented: we expect interventions to produce much the same results in the future as they did in the past, and this makes planning possible.

The evidence for general causal claims is thought to emanate from RCTs, so let us turn again to RCTs to see whether they supply us with evidence that can support such claims. It would seem that we largely assume that they do. The Department of Education's guide, as we have seen, presupposes that two RCTs are sufficient to demonstrate general effectiveness. Keith Morrison (2001) thinks that advocates of EBP simply assume that RCTs ensure generalizability, which is, of course, exactly what one wants in EBP—if results are generalizable, we may assume that the effect travels to other target populations so that results are reproducible and we can plan for their attainment. But does RCT evidence tell us that a cause holds widely? No, Cartwright (2007a) argued, RCTs require strong premises, and strong premises do not hold widely. Because of design restrictions, RCT results hold formally for the study group (the sample) and only for that group, she insists. Methods that are strong on internal validity are correspondingly weak on external validity. RCTs establish efficacy, not effectiveness. We tend to assume without question, Cartwright argues, that efficacy is evidence for effectiveness. But we should not take this for granted—either it presumes that the effect depends exclusively on the intervention and not on who receives it, or it relies on presumed commonalities between the study group and the target group. This is a matter of concern to EBP and its advocates, because if Cartwright is correct, RCT evidence does not tell us what we think it tells us. Multiple RCTs will not solve this problem; the weakness of enumerative induction—inferences from single instances to a general conclusion—is well known. So how then can we ground our expectation that results are reproducible and can be planned for?

The Practitioner Perspective

EBP, as it is mostly discussed, is researcher centered. The typical advice guides, such as that of the What Works Clearinghouse, tend to focus on the finding of causes and the quality of the evidence produced. Claims and interventions should be rigorously tested by stringent methods such as RCTs and ranked accordingly. The narrowness of the kind of evidence thus admitted (or preferred) is pointed out by many critics, but it is of equal importance that the kind of claims RCT evidence is evidence for is also rather narrow. Shifting the focus from research to practice significantly changes the game. And bring in the practitioners we must—EBP is eminently practical in nature, concerning as it does the production of desirable results. Putting practice center stage means shifting from *finding* causes and assessing the quality of research evidence to *using* causes to produce change. In research we can control for confounders and keep variables fixed. In practice we can do no such thing; hence the significant change of the game.

The claim a practitioner wants evidence for is not the same claim that a researcher wants evidence for. The researcher wants evidence for a causal hypothesis, which we have seen can be of many different kinds, for example, the contribution of X to Y. The practitioner wants evidence for a different kind of claim—namely, whether X will contribute positively to Y for his students, in his context. This is the practitioner's problem: the evidence that research provides, rigorous as it may be, does not tell him whether a proposed intervention will work *here*, for this particular target group. Something more is required.

Fidelity

Fidelity is a demand for faithfulness in implementation: if you are to implement an intervention that is backed by, say, two solid RCTs, you should do it exactly as it was done where the evidence was collected. The minimal definition of EBP adopted here leaves it open whether fidelity should be included or not, but there can be no doubt that both advocates and critics take it that it is—making fidelity one of the most controversial issues in EBP. The advocate argument centers on quality of implementation (e.g., Arnesen, Ogden, & Sørli, 2006). It basically says that if X is implemented differently than is prescribed by researchers or program developers, we can no longer know exactly what it is that works. If unfaithfully implemented, the intervention might not produce the expected results, and the program developers cannot be held responsible for the results that do obtain. Failure to obtain the expected results is to be blamed on unsystematic or unfaithful implementation of a program, the argument goes. Note that the results are described as expected.

The critics, on the other hand, interpret fidelity as an attempt to curb the judgment and practical knowledge of the teachers; perhaps even as an attempt to replace professional judgment with research evidence. Biesta (2007), for example, argues that in the EBP framework the only thing that remains for practitioners to do is to follow rules for action. These rules are thought to be somehow directly derived from the evidence. Biesta is by no means the only EBP critic to voice this criticism; we find the same view in Bridges, Smeyers, and Smith (2008):

The evidence-based policy movement seems almost to presuppose an algorithm which will generate policy decisions: *If A is what you want to achieve and if research shows R1, R2 and R3 to be the case, and if furthermore research shows that doing P is positively correlated with A, then it follows that P is what you need to do.* So provided you have your educational/political goals sorted out, all

you need to do is slot in the appropriate research findings—the right information—to extract your policy. (p. 9)

No consideration of the concrete situation is deemed necessary, and professional judgment therefore becomes practically superfluous. Many critics of EBP make the same point: teaching should not be a matter of following rules, but a matter of making judgments. If fidelity implies following highly scripted lessons to the letter, the critics have a good point. If fidelity means being faithful to higher level principles, such as “provide feedback on home assignments,” it becomes more open and it is no longer clear exactly what one is supposed to be faithful to, since feedback can be given in a number of ways. We should also note here that EBP advocates, for example, David Hargreaves (1996b), emphatically insist that evidence should enhance professional judgment, not replace it. Let us also note briefly the usage of the term “evidence,” since it deviates from the epistemological usage of the term. Biesta (and other critics) picture evidence as something from which rules for action can be inferred. But evidence is (quantitative) data that speak to the truth value of a causal hypothesis, not something from which you derive rules for action. Indeed, the word “based” in evidence-based practice is misleading—practice is not based on the RCT evidence; it is based on the hypothesis (supposedly) supported by the evidence. Remember that the role of evidence can be summed up as *support*. Evidence surely can enhance judgment, although EBP advocates tend to be rather hazy about how this is supposed to happen, especially if they also endorse the principle of fidelity.

Contextual Matters

If we hold that causes are easily exportable and can be relied on to produce their effect across a variety of different contexts, we rely on a number of assumptions about causality and about

contexts. For example, we must assume that the causal X–Y relation is somehow basic, that it simply holds in and of itself. This assumption is easy to form; if we have conducted an RCT (or several, and pooled the results in a meta-analysis) and found a relation between an intervention and an effect of a decent magnitude, chances are that we conclude that this relation simply exists. Causal relations that hold in and of themselves naturally also hold widely; they are stable, and the cause can be relied on as sufficient to bring about its effect most of the time, in most contexts, if not all. This is a very powerful set of assumptions indeed—it underpins the belief that desirable results are reproducible and can be planned for, which is exactly what not only EBP wants but what practical pedagogy wants and what everyday life in general runs on.

The second set of assumptions concerns context. The US Department of Education guide (2003) advises that RCTs should demonstrate the intervention’s effectiveness in school settings similar to yours, before you can be confident that it will work for you. The guide provides no information about what features should be similar or how similar those features should be; still, a common enough assumption is hinted at here: if two contexts are (sufficiently) similar (on the right kind of features) the cause that worked in one will also work in the other. But as all teachers know, students are different, teachers are different, parents are different, headmasters are different, and school cultures are different. The problem faced by EBP is how deep these differences are and what they imply for the exportability of interventions.

On the view taken here, causal relations are not general, not basic, and therefore do not hold in and of themselves. Causal relations are context dependent, and contexts should be expected to be different, just as people are different. This view poses problems for the practitioner, because it means that an intervention that is shown by an RCT to work *somewhere* (or in many *somewheres*) cannot simply be assumed to work *here*. Using causes in practice to bring about

desirable changes is very different from *finding* them, and context is all-important (Cartwright, 2012).

All interventions are inserted into an already existing practice, and all practices are highly complex causal/social systems with many factors, causes, effects, persons, beliefs, values, interactions and relations. This system already produces an output Y; we are just not happy with it and wish to improve it. Suppose that most of our first graders do learn to read, but that some are reading delayed. We wish to change that, so we consider whether to implement Hatcher's method. We intervene by changing the cause that we hold to be (mainly) responsible for Y—namely, X—or we implement a brand-new X. But when we implement X or change it from x_i to x_j (shifting from one method of reading instruction to another), we generally thereby also change other factors in the system (context, practice), not just the ones causally downstream from X. We might (inadvertently) have changed both A, B, and C—all of which may have an effect on Y. Some of these contextual changes might reinforce the effect of X; others might counteract it. For example, in selecting the group of reading-delayed children for special treatment, we might find that we change the interactional patterns in the class, and that we change the attitudes of parents toward their children's education and toward the teacher or the school. With the changes to A, B, and C, we are no longer in system_g but in system_h. The probability of Y might thereby change; it might increase or it might decrease. Hence, insofar as EBP focuses exclusively on the X–Y relation, natural as this is, it tells only half the story. If we take the context into account, it transpires that if X is going to be an efficacious strategy for changing (bringing about, enhancing, improving, preventing, reducing) Y, then it is not the relation between X and Y that matters the most. What matters instead is that the probability of Y given X-in-conjunction-with-

system is higher than the probability of Y given not-X-in-conjunction-with-system. But what do we need to know to make such judgments?

Relevance and Evidence

On the understanding of EBP advanced here, fidelity is misguided. It rests on causal assumptions that are at least problematic; it fails to distinguish between finding causes and using causes; and it fails to pay proper attention to contextual matters.

What, then, should a practitioner look for when trying to make a decision about whether to implement X or not? X has worked *somewhere*; that has been established by RCTs. But when is the fact that X has worked somewhere relevant to a judgment that X will also work *here*? If the world is diverse, we cannot simply export a causal connection, insert it into a different context, and expect it to work there. The practitioner will need to gather a lot of heterogeneous evidence, put it together, and make an astute all-things-considered judgment about the likelihood that X will bring about the desired results *here* were it to be implemented. The success of EBP depends not only on rigorous research evidence but also on the steps taken to use an intervention to bring about desirable changes in a context where the intervention is as yet untried.

What are the things to be considered for an all-things-considered decision about implementing X? First, the practitioner already knows that X has worked *somewhere*; the RCT evidence tells him or her that. Thus, we do know that X played a positive causal role for many of the individuals in the study group (but not necessarily all of them; effect sizes are aggregate results and thus compatible with negative results for some individuals).

Second, the practitioner must think about how the intervention might work if it were implemented. RCTs run on an input–output logic and do not tell us anything about *how* the cause

is thought to bring about its effect. But a practitioner needs to ask whether X can play a positive causal role in his or her context, and then the question to ask is *how*, rather than *what*.

Third, given our understanding of causes as INUS conditions, the practitioner will have to map the contextual factors that are necessary for X to be able to do its work and bring about Y. What are the enabling factors? If they are not present, can they be easily procured? Do they outweigh any disabling factors that may be present? It is important to remember that enablers may be absences of hindrances. Despite their adherence to the principle of fidelity, Arnesen, Ogden, and Sørli (2006) acknowledge the importance of context for bringing about Y. For example, they point out that there must be no staff conflicts if the behavioural program is to work. Such conflicts would be a contextual disabler, and their absence is necessary. If you wish to implement Hatcher's method, you have to look at your students and decide whether you think this will suit them, whether they are motivated, and how they might interact with the method and the materials. As David Olson (2004) points out, the effect of an intervention depends on how it is "taken" or understood by the learner. But vital contextual factors also include mundane things such as availability of adequate materials, whether the parents will support and help if the method requires homework, whether you have a suitable classroom and sufficient extra time, whether a teacher assistant is available, and so on. Hatcher's method is the INUS condition, the salient factor, but it requires a contextual support team to be able to do its work.

Fourth, the practitioner needs to have some idea of how the context might change as a result of implementing X. Will it change the interactions among the students? Create jealousy? Take resources meant for other activities? The stability of the system into which an intervention is inserted is generally of vital importance for our chances of success. If the system is shifting and unstable X may never be able to make its effect happen. The practitioner must therefore know

what the stabilizing factors are and how to control them (assuming they are within his or her control).

In sum, the INUS approach to causality and the all-important role of contextual factors and the target group members themselves in bringing about results strongly suggest that fidelity is misguided. The intervention is not solely responsible for the result; one has to take both the target group (whatever the scope) and contextual factors into consideration. On the other hand, similarity of contexts loses its significance because an intervention that worked somewhere can be made to be relevant here—there is no reason to assume that one needs exactly the same contextual support factors. The enablers that made X work *there* need not the same enablers that will make X work *here*. What is important is that the practitioner carefully considers *how* X can be made to work in his or her context.

Conclusion

EBP is a complex enterprise. The seemingly simple question of using the best available evidence to bring about desirable results and prevent undesirable ones branches out in different directions to involve problems concerning what educational research can and should contribute to practice, the nature of teaching, what kind of knowledge teachers need, what education should be all about, how we judge what works, the role of context and the exportability of interventions, what we think causality is, and so on. We thus meet both ontological, epistemological, and normative questions.

It is important to distinguish between the evidence and the claim which it is evidence *for*. Evidence serves to support (confirm, disconfirm) a claim, and strictly speaking practice is based

on claims, not on evidence. Research evidence (as well as everyday types of evidence) should always be evaluated for its trustworthiness, its relevance, and its scope.

EBP as it is generally discussed emphasizes research at the expense of practice. The demands of rigor made on research evidence are very high. There is a growing literature on implementation and a growing understanding of the importance of quality of implementation, but insofar as this focuses on fidelity, it is misguided. Fidelity fails to take into account the diversity of the world and the importance of the context into which an intervention is to be inserted. It is argued here that implementation centers on the matter of whether an intervention will work *here* and that a reasonable answer to that question requires much local, heterogeneous evidence. The local evidence concerning target group and context must be provided by the practitioner. The research evidence tells only part of the story.

If EBP is to be a success, the research story and the local-practice story must be brought together, and this is the practitioner's job. The researcher does not know what is relevant in the concrete context faced by the practitioner; that is for the practitioner to decide.

EBP thus demands much knowledge, good thinking, and astute judgments by practitioners.

As a recommendation for future research, I would suggest inquiries into how the research story and the contextual story come together; how practitioners understand the causal systems they work within, how they understand effectiveness, and how they adapt or translate generalized guidelines into concrete local practice.

References

Achinstein, P. (2001). *The book of evidence*. Oxford: Oxford University Press.

Arnesen, A., Ogden, T., & Sørli, M.-A. (2006). *Positiv atferd og støttende læringsmiljø i skolen*. Oslo: Universitetsforlaget.

Biesta, G. (2007). Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, 57, 1–22.

Biesta, G. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Boulder, CO: Paradigm.

Bridges, D., Smeyers, P., & Smith, R. (2008). Educational research and the practical judgment of policy makers. *Journal of Philosophy of Education*, 42(Suppl. 1), 5–11.

Cartwright, N. (2007a). Are RCTs the gold standard? *BioSocieties*, 2, 11–20.

Cartwright, N. (2007b). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge, U.K.: Cambridge University Press.

Cartwright, N. (2012). Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, 79, 973–989.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press.

Davies, P. (2004). Systematic reviews and the Campbell Collaboration. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 21–33). Maidenhead, U.K.: Open University Press.

Dean, C. B., Hubbell, E. R., Pitler, H., & Stone, B. J. (2012). *Classroom instruction that works: Research-based strategies for increasing student achievement* (2d ed.). Denver, CO: Mid-continent Research for Education and Learning.

Hammersley, M. (1997). Educational research and teaching: A response to David Hargreaves’ TTA lecture. *British Educational Research Journal*, 23, 141–161.

Hammersley, M. (2004). Some questions about evidence-based practice in education. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 133–149).

Maidenhead, U.K.: Open University Press.

Hargreaves, D. (1996a). Educational research and evidence-based educational practice: A response to critics. *Research Intelligence*, 58, 12–16.

Hargreaves, D. (1996b). Teaching as a research-based profession: Possibilities and prospects.

Teacher Training Agency Annual Lecture, London. Retrieved from

<https://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/TTA%20Hargreaves%20lecture.pdf>.

Hatcher, P., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., Gibbs, S., . . . Snowling, M. J.

(2006). Efficacy of small group reading intervention for readers with reading delay: A randomised controlled trial. *Journal of Child Psychology and Psychiatry*, 47(8), 820–827.

Hitchcock, D. (2011). Instrumental rationality. In P. McBurney, I. Rahwan, & S. Parsons

(Eds.), *Argumentation in multi-agent systems: Proceedings of the 7th international ArgMAS Workshop* (pp. 1–11). New York: Springer.

Kelly, T. (2008). Evidence. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Retrieved

from <http://plato.stanford.edu/entries/evidence/>.

Kvernbekk, T. (2016). *Evidence-based practice in education: Functions of evidence and causal presuppositions*. London: Routledge.

Lærerorganisasjonenes skolenevnd. (1933). *Innstilling*. Oslo: O. Fredr. Arnesens Bok- og Akcidenstrykkeri.

Mackie, J. L. (1975). Causes and conditions. In E. Sosa (Ed.), *Causation and conditionals* (pp.

15–38). Oxford: Oxford University Press.

- Menzies, P., & Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science*, 44, 187–203.
- Morrison, K. (2001). Randomised controlled trials for evidence-based education: Some problems in judging “what works.” *Evaluation and Research in Education*, 15(2), 69–83.
- Oancea, A., & Pring, R. (2008). The importance of being thorough: On systematic accumulation of “what works” in education research. *Journal of Philosophy of Education*, 42(Suppl. 1), 15–39.
- Olson, D. R. (2004). The triumph of hope over experience in the search for “what works”: A response to Slavin. *Educational Researcher*, 33, 24–26.
- Pawson, R. (2012). *Evidence-based policy: A realist perspective*. Los Angeles: SAGE.
- Schaffer, J. (2007). The metaphysics of causation. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Retrieved from <http://plato.stanford.edu/entries/causation-metaphysics/>.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31, 15–21.
- Slavin, R. E. (2004). Education research can and must address “what works” questions. *Educational Researcher*, 33, 27–28.
- U.S. Department of Education. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: Coalition for Evidence-Based Policy. Retrieved from <http://www2.ed.gov/rschstat/research/pubs/rigorousetid/rigorousetid/pdf>.
- Worrall, J. (2007). Why there’s no cause to randomize. *British Journal for the Philosophy of Science*, 58, 451–488.

Bibliography

Biesta, G. (2010). Why “what works” still won’t work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, 29, 491–503.

Bridges, D., & Watts, M. (2008). Educational research and policy: Epistemological considerations. *Journal of Philosophy of Education*, 42(Suppl. 1), 41–62.

Cartwright, N. (2013). Knowing what we are talking about: Why evidence doesn’t always travel. *Evidence and Policy*, 9, 97–112.

Cartwright, N., & Munro, E. (2010): The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16, 260–266.

Cartwright, N., & Stegenga, J. (2011). A theory of evidence for evidence-based policy. *Proceedings of the British Academy*, 171, 289–319.

Hammersley, M. (Ed.). (2007). *Educational research and evidence-based practice*. Los Angeles: SAGE.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. London: Routledge.

Kvernbekk, T. (2013). Evidence-based practice: On the functions of evidence in practical reasoning. *Studier i Pædagogisk Filosofi*, 2(2), 19–33.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge, U.K.: Cambridge University Press.

Phillips, D. C. (2007). Adding complexity: Philosophical perspectives on the relationship between evidence and policy. In P. Moss (Ed.), *Evidence and decision making*. Yearbook of

the National Society for the Study of Education, 106 (pp. 376–402). Malden, MA: Blackwell.

Psillos, S. (2009). Regularity theories. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 131–157). Oxford: Oxford University Press.

Reiss, J. (2009). Causation in the social sciences: Evidence, inference, and purpose. *Philosophy of the Social Sciences*, 39, 20–40.

Rosenfield, S., & Berninger, V. (Eds.). (2009). *Implementing evidence-based academic interventions in school settings*. Oxford: Oxford University Press.

Sanderson, I. (2003). Is it “what works” that matters? Evaluation and evidence-based policy-making. *Research Papers in Education*, 18, 331–345.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. Oxford: Oxford University Press.

Smeyers, P., & Depaepe, M. (Eds.). (2006). *Educational research: Why “what works” doesn’t work*. Dordrecht, The Netherlands: Springer.

Thomas, G., & Pring, R. (Eds.). (2004). *Evidence-based practice in education*. Maidenhead, U.K.: Open University Press.

Williamson, J. (2009). Probabilistic theories. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 185–212). Oxford: Oxford University Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Tone Kvernbekk