ISSA Proceedings 2014 – Evidence-Based Practice: Evidence Set In An Argument

Abstract: Evidence-based practice (EBP) is currently a dominating trend in many professional areas. But what do we want evidence for in EBP? Evidence generally speaks to the trustworthiness of our beliefs, but EBP is practical in nature and truth is not really what is at stake. Rather we are after effectiveness in bringing about changes. What we need evidence *for* is a prediction to the effect that what has worked in one context will also work *here*. In this paper I argue that is makes good sense to view this prediction as the conclusion of an argument. To set the evidence in an argument will structure our thinking and help us focus on what kinds of evidence we need to support the likelihood that an intervention *here* will work.

Keywords: Argument, causal role, EBP, effectiveness, enablers, evidence, external validity, local facts, RCT, stability of context

1. *Introduction*

There exists a vast literature on EBP, hardly surprising given the status of 'evidence-based' as a buzzword in contemporary professional debates, such as education, medicine, psychiatry and social policy. Researchers are responding in many ways to political demands for better research bases to inform and guide both policy and practice; some by producing the kind of evidence it is assumed can serve as a base for practice; others by criticizing or even rejecting the whole enterprise of EBP – the latter frequently, but not exclusively, couched in terms of worries about instrumentalization of practice and restrictions in the freedom of professionals to exercise their judgment.

EBP is practical in nature. It is commonly called the *what works* agenda and its focus is the use of the best available evidence in the bringing about of desirable goals, both for client and society. This is indeed my preferred minimal definition of EBP; the production of desirable change, or conversely how we intervene to prevent certain undesirable outcomes. It is vital to note that EBP is deeply causal: we intervene into a "system" which already produces an output in order to

change that output in a desirable direction. These interventions should be based on evidence that shows what works. To say that something (an intervention of some kind) works, is to say that doing it brings us the effects we want. For short, do X and it will lead to Y.

The very term 'evidence-based practice' obviously draws attention to evidence. Generally, epistemologists seem to agree that the term 'evidence' denotes that which serves to confirm or disconfirm a theory (claim, belief, hypothesis) (e.g. Achinstein, 2001; Kelly 2008). The basic function of evidence is thus summed up in the word *support*. Evidence speaks to the truth value and the trustworthiness of a claim, and is therefore relevant to all belief formation processes, whether in research or in daily life, including the ones where we form beliefs about the causal relation between action and result, input and output. This basic function can, I submit, in principle be performed by all sorts of data, facts and personal experiences. Indeed, it is worth pointing out that all people have first-hand experiences of the causal kind that we are talking about here. To act as an agent means to intervene in the world and have an influence on it (Menzies and Price, 1993). At the same time, 'evidence-based practice' has led to many misunderstandings about the role of evidence as well as to the crux of the matter being overlooked. What is really at stake is the *claim* that the evidence is evidence for. Evidence is in a sense a servant; good evidence provides us with good reason to believe that the claim is true.

I shall in this paper argue that setting evidence is an argument makes good sense for the practical enterprise of EBP; it serves to clarify and structure our thinking about what we need to know. But to see that, we first have to look at the basic causal structure of EBP and the EBP orthodoxy concerning admissible forms of evidence as well as assumptions concerning uses of evidence. Thus, this paper is mainly a laying-out of the premises I suggest are needed to bolster the conclusion that EBP will be well served by setting the evidence in an argument.

2. *The causal nature of EBP*
The short version of the causal nature is that EBP is causal because it is about the *bringing about* of desirable results. That is to say, we have a causal connection between an action or intervention and its effects; between X and Y. The long version of the causal nature of EBP takes into account the many forms of causation; direct, indirect, necessary, sufficient, probable, generic, actual, etc. and develops a more complex and sophisticated picture. In educational contexts,

as, I assume, in political contexts, this causal complexity goes highly unrecognized. However, for my purposes in this paper a simplified X-Y relation will by and large do.

My own field is education; a complex field with many factors that interact and influence each other in many different ways. Interventions also vary in nature, from simple actions to highly complex school-wide projects which may take two or three years to run. It is essential to be aware that, regardless of field, any intervention is inserted into pre-existing conditions. The causal system into which we intervene already produces an output; we just wish to change it because we are not entirely happy with the output – in education, student achievement is a typical output of this sort. The already existing output is termed the *default value* (Hitchcock, 2007, p.506); the value we would expect a variable such as student achievement to have in the absence of intervening causes. The default assumption is that the system will persist in its state and keep producing the default results unless we do something or something happens. The default, Hitchcock stresses, it not that the state or value in question *is* this or that, but that it will remain this or that unless something happens to change it. When a set of variables all take on their default value and business is run as usual, they cannot by themselves take on a different value. This is a natural principle of causal reasoning, Hitchcock thinks. We tend to assume that if a variable should take on a deviant (or unexpected) value, there must be some outside variable or event that explains it. That is, to change the value of our target variable, whether student achievement or some other desirable outcome, we have to intervene somehow. This certainly seems to be a tacit presupposition of EBP.

For various reasons, the causal theory that best suits the logic of EBP is the manipulationist theory of causation (e.g. Pearl, 2009; Sloman, 2005; Woodward, 2003, 2008). Let us suppose that X produces Y as its default result. To change the value of Y, we must change the value of X. Thus, if we set the value of X to *xi* rather than *xk*, then the value of Y should follow in train and change to *yi* rather than *yk*. This is precisely what the manipulationist theory of causation tells us: there is an intimate connection between causation and manipulation such that causal relationships are eminently exploitable for the purpose of change. This is one of the reasons why this theory of causation is so popular in disciplines which are to bring about change and development as well as give recommendations for practice and policy.

The point of intervening is that we set the value of X to *xi* from outside the system rather than letting X be decided by the other variables in the system. That is to say, we manipulate X in order to further the changes in Y we deem desirable, naturally on the assumption that X actually leads to or brings about Y. As Judea Pearl puts it,

T*he simplest type of external intervention is one in which a single variable, say Xi, is forced to take on some fixed value xi. Such an intervention, which we call "atomic," amounts to lifting Xi from the old functional mechanism xi = fi(pai, ui) and placing it under the influence of a new mechanism that sets the value xi while keeping all other mechanisms unperturbed* (2009, p. 70).

There is, however, more to intervention than this quote tells us. First, it changes the value of Y, even though this is not explicitly mentioned in Pearl's definition. Changing Y is the main aim of educational interventions and usually the reason why we intervene in the first place (e.g. to improve student achievement). Second, the intervention changes the entire causal model because it cuts the effect (*yk*) off from its normal causes (*xk*). When we have intervened on X, the system no longer continues in its default state. Business is no longer run as usual, but is now running in a different way, one we think (or hope) should bring about the desired result or at least increase its probability. Third, the intervention disrupts the relationship between X and its parents. The value of X is no longer determined by the default running of the system, but by the intervention. All other influences on X have been blocked and/or cut off. As the equation in the quote indicates, X*i* is lifted from the influence of P, its parents, and U, an error term representing the impact of omitted and/or unknown variables, and its value is decided by a new mechanism, namely the intervention. I prefer to interpret this in line with the causal agency advocated by Menzies and Price, although Pearl himself states that intervention does not necessarily have to involve human activity. But in education interventions require agency, hence my adoption of Menzies and Price's view on this point.

This is not the place to discuss manipulationist theory in detail, but a couple of issues deserve mention. First, there is Pearl's view that causal mechanisms (X-Y relations) are autonomous. He thus argues that our intervention on one causal connection leaves the other connections in the system undisturbed. This presupposition seems deeply problematic to me. Educational practice is best understood as an open system where events, actions and factors are somehow

locked together, obviously to varying degrees. If factors hang together, the change in Y will depend more on the total structure and it is a mistake – however tempting it is – to look at only small chunks or individual causal mechanisms. In complex systems we cannot assume that intervention on one mechanism leaves all other mechanisms intact. Second, it seems to be a presupposition of manpulationist theorists that X is already a part of the system. For example, Christopher Hitchcock (2007) argues that X-Y relations are internal to the system and that interventions therefore involve exogenous changes to X. My point here is twofold. Firstly, in education a teacher, as an agent *within* the system can decide to make changes in input X; this qualifies as an intervention in the broad sense of them term, but it comes from inside the system and is thus not exogenous. Second, there are many EBP cases where X is exogenous and inserted into the system as a new element. I view these two points as unproblematic amendments the manipulationist theory of causation. The main point is that X be manipulable and that the intervention alters the causal system.

It is the ambition of EBP to provide knowledge that *works*; that is, to provide knowledge about how causal input X can be changed to produce desired changes in output Y. For example how implementation of a reading instruction program can improve the reading skills of slow readers, or how a school-wide behavioural support program can serve to enhance students' social skills and prevent future problem behaviour. But not only that – we wish to know what works generally. That means not only that the effect (output, result) in question is reproducible in principle, but that we know how to achieve it regularly and can plan for it. This kind of practical causal knowledge is future-oriented, in the sense that we, on the basis of experience or other empirical evidence, form the expectation that the desirable results obtained somewhere can somehow be reproduced.

3. *What does the evidence tell us?*
As suggested above, the basic function of evidence is to speak to the truth value of beliefs. In the EBP case, both advocates and critics simply assume that the evidence speaks to the truth of the belief that there is a causal connection between X and Y, and that this is all the evidence there is (or all we need).

In a similar manner, both advocates and critics often understand EBP to include a hierarchy of evidence as part of its definition. There are various versions of this hierarchy; what they have in common is that they all rank randomized controlled trials (RCTs) on top, and that professional judgment is ranked at or near the

bottom (see e.g. Pawson, 2012). The standard criticism is that such hierarchies unduly privilege certain forms of knowledge and research designs, undervalue the contributions of other research perspectives, and especially that they undervalue professional experience and judgment. The privileging of RCT evidence is evident in e.g. the US Department of Education's *User Friendly Guide*. EBP literature, such as the *User Friendly Guide*, provides evidence-ranking schemes (which tell us that the best evidence comes from RCTs), it provides advice guides (which tell us to choose an educational intervention that is backed by good (RCT) evidence, and it often provides "warehouses" (where we find interventions backed by good evidence). Together these three different functions make up the foundation of what has become known as the EBP orthodoxy (see e.g. Cartwright and Hardie, 2012). There is another element to the orthodoxy that I shall return to below.

There are good reasons to adopt the EBP orthodoxy and even better reasons not to adopt it. The principle behind evidential ranking schemes is *trustworthiness* – our evidence needs to be trustworthy or reliable in order to do its job, which is to speak to the truth value of claims and beliefs. It is no accident that RCTs have established themselves as the gold standard. Nancy Cartwright (2007) divides all research methods in two; *clinchers* and *vouchers.* RCTs are clinchers: methods that are deductive and whose logic is such that if all the specific assumptions of the trial are met, a positive result will logically entail the conclusion. The evidence provided is thus sufficient for the conclusion; one might even say that it guarantees it. The evidence, in turn, is guaranteed by the research design. In RCTs we compare groups that are the same with respect to all relevant (causal) factors except one. Random assignment is supposed to ensure that the groups have the same distribution of causal and other factors. The standard result of an RCT is a "treatment effect" (expressed in terms of effect size): average effect in treatment group minus average effect in control group. We assume that the difference between the two groups needs a causal explanation, and since other factors (supposedly) are equally distributed we infer that the treatment, our intervention, is the cause of the outcome. It works; we might be tempted to conclude.

RCTs are strong on internal validity. If we obtain an average positive result and the conditions of the trial are met, we may safely conclude that the causal claim in question is true, X does indeed bring about Y and the evidence supports it. But internal validity is purchased at the expense of external validity, or generality. As

Nancy Cartwright (2007) argues, what RCT evidence shows is strictly speaking that the X-Y relation holds where the trial was conducted, for that particular study group (see Cartwright for a detailed discussion of the limitations of the research design). It by no means shows that the X-Y relation holds generally across differing contexts. This fact is not discussed in the EBP literature. Rather, we seem to take it for granted that RCT evidence shows that the causal X-Y relation holds in general, that something works in general. The fact that it does not, is a major premise in the argument for why it is important to set evidence in an argument.

There are several sides to the limitation of RCT evidence. First, we here come across a problem that is also found in the manipulationist theory of causation; namely that one does not distinguish between *finding* and using causes. Manipulationist theory and empirical research designs alike focus on finding causes. To investigate whether X causes Y, we see if the two are correlated once we have controlled for other possible causes of Y. We hold various background factors fixed, manipulate the values of X and observe whether the values of Y change in train. Basically we conclude that X causes Y if the probability of Y is higher with X than without it, and the evidence we get supports our view. But *using* causes to bring about desired changes is another matter altogether. I am tempted to say that both manipulationist theory, RCTs and EBP only tell us half the story. They all think in terms of methodology geared at finding causes. When it comes to using causes, it is not the relation between X and Y that matters the most. When we implement an intervention we either change an X that is already part of the system, or we insert it into the system. Either way the pre-existing system, *practice*, has to be taken into account when we use causes. Hence, what matters is that the probability of Y given X-in-conjunction-with-systemi is larger than the probability of Y given not-X-in-conjunction-with-systemi. And the RCT tells us nothing about this. As Cartwright (2009) points out, the formula that shows that X is a cause of Y, for example expressed in terms of a treatment effect, need not be the right formula for telling whether X will produce Y when we implement it in some concrete system. When we implement X, we generally also change other factors in the system, not only the ones causally downstream from X. But the RCT evidence does not tell us whether X will also affect A, B, and C, and if so how that will affect Y.

The second ramification of the limitation of RCT evidence is a corollary of the

first, and concerns the EBP orthodoxy. This orthodoxy also demands faithfulness in implementation, termed fidelity. If you are to implement in your context an intervention that an RCT tells you has worked somewhere, you should do it exactly as it was done there. Take for example a school-wide behavioural program (Arnesen, Ogden and Sørlie, 2006). Components, principles and guidelines are decided in advance, and so is their order and manner of implementation, although the authors concede that some local adjustment is necessary. But basically implementers must be loyal to the procedures prescribed by the program developers. If actual implementation deviates from prescribed implementation, we no longer know exactly *what* it is that works, the argument goes, and the program suppliers cannot be held responsible for the results. Variations in the efficacy of X are generally due to deviant or unsystematic implementation, the EBP orthodoxy holds. The orthodoxy presupposes similarity of contexts and generality of X-Y relation. The demand for fidelity in EBP is misguided, as it tacitly assumes that the RCT evidence showing the effect of X on Y is all you need.

But what do practitioners need evidence *for?* I propose that what practitioners, say teachers, want evidence for, is a prediction that X will work *here*, in my classroom, were I to implement it. The RCT evidence only speaks indirectly to that question, by telling you that X worked *somewhere.* But how do you get from *somewhere* to *here*? This is where the usefulness of an argument comes in.

4. *Setting evidence in an argument*
Let me back up a little. It is important that we take on board the fact that contributions to an outcome both can and generally do come from different sources. This sounds commonplace, but is easily forgotten; we tend to look for *the* cause and if we implement an intervention it is only natural that this intervention is salient for us and we ignore other factors. But the overall effect on Y depends on how all these factors add up; thus, an intervention is part of a team of causes and enabling factors which work together.

What, then, should a practitioner look for when trying to make a decision about whether to implement X or not? Which facts must be collected if I am to hedge my bets that X will work *here?* When is the fact that X worked *there* relevant to the prediction that it will also work *here?* We cannot take it for granted that it will, no matter how large the effect size emanating from the RCT evidence. We cannot simply export a causal connection and insert it into a different context and expect it to work. Causal principles are local, Cartwright argues, and it is easy to agree

with her. Educational practitioners love to point out that students are different, teachers are different, curricula are different, headmasters are different, parents are different, and school cultures are different. So how can the RCT evidence be made relevant?

I assume that what practitioners want to know is whether an intervention is worth trying in their own concrete context. Will X work here, that is, make a positive causal contribution here if I implement it? RCT evidence does not tell them that. What is does tell them, is that X made a positive contribution to Y somewhere, *and* that given this positive contribution, we may infer that certain enabling factors were present which allowed X to do its work and make its way to Y. That is to say, the other factors necessary for producing the outcome must also be in place – it is vital to remember that our intervention is part of a constellation of causes which together bring about Y. An effectiveness prediction that X will work here must take the whole constellation into account, as well as possible. It is this task that is made easier and more systematic by thinking of the effectiveness prediction as the conclusion of an argument and that the job is to gather the premises which lead up to the conclusion.

What works *somewhere*, as shown by the RCT evidence, can be made relevant to what will work here. But a number of other facts must be collected if we are to say something about X-in-conjunction-with-system, which is what we want:

* In "our" context *here* we already get an outcome, a default result, concerning the student achievements in question, but we want to improve them. How are these results produced? What factors are present in our context and how do they combine to produce the result?

* This constellation of causes is called the causal principle for the outcome (Cartwright and Hardie, 2012 and it is needed to connect the alleged cause with the desired effect.

* Mapping the local causal principle is not enough. Next we have to look at the proposed intervention X and ask whether it can play a positive causal role for producing the desired effect in our setting. How can it work? There is no substitute, Cartwright and Hardie insist, for thinking thoroughly about how X might work if implemented.

* Next we look at the factors that must be in place if the intervention is to be able

to play its causal role. Which are they? Are they present? If not present, can they be easily procured? Do they outweigh any disabling factors that might be in place? It is important to remember that some of these enablers may be absences of hindrances. Arnesen, Ogden and Sørlie (2006) provide examples of such local facts, despite their adherence to the EBP orthodoxy and the principle of fidelity. For example, they argue that there must not be personal conflicts among the staff if the behavioural program is to work positively. That is, a conflict is a contextual disabler which hinders or obstructs the working of the program. Conversely we might say that its absence is an enabler. Another local enabling factor is the fact that staff norms and values at least do not contradict the values inherent in the program to be implemented.

* Not only must the necessary enablers be in place, their organization must also be stable. The stability of the system into which we contemplate inserting X is of vital importance for our chances of success. If the system is shifting and unstable, X may never be able to do its work and produce Y. This fact is well-known to teachers, but perhaps not really recognized by EBP proponents. But teachers seek to stabilize the environment, by structuring it in different ways: creating and enforcing rules of conduct, establishing habits and ways of doing things – in short, creating a stable environment which at least to some degree makes for predictability and thus allows us to expect with some confidence that our plans will work out. Time-honoured educational domains such as curriculum theory and didactics can be viewed in this light: they provide knowledge and advice on how to create the stable conditions necessary for goal achievement in general. But since we are trying to predict whether X will work if we implement it, the stability conditions we assess must be linked to X.

* It should be noticed that this kind of "mapping" is not about listing similarities between *somewhere* and *here*. Similarities are not important for this kind of generalization. Rather, what it takes is that we have some idea about what a good constellation of factors surrounding X might be, factors which enable X to make a positive contribution to Y. This constellation need not be the same; it can vary from context to context. The important thing is that we map the enablers, procure them if necessary, and that we avoid or remove the disablers.

In sum: local facts are as necessary as they are overlooked. I by no means claim that the issues listed above comprise an exhaustive list of facts a practitioner needs to map in order to hedge his or her bets that a given intervention will work

should it be implemented *here*. Yet it should be evident from this set of issues that it takes a lot of deliberation to figure out the chances that an intervention might work. Setting all these different kinds of evidences into a (reasonably) clear argument structure helps us sort them out and see what facts we need to ascertain. Inspired by Cartwright and Hardie, here is what I propose:

Premise 1: The intervention in question, X, worked *somewhere*; that is, it played a positive causal role in achieving Y for at least some of the individuals in the study group. The RCT evidence tells us that, and it also indicates how strong the causal influence of X on Y is, given that all other factors are held fixed (the effect size). We should remember, however, that effect size is a statistical entity and only informs us of the aggregate result. A positive aggregate result is perfectly compatible with negative results for some of the individuals in the study group.

Premise 2: Which factors govern the default production of Y *here*? The RCT evidence does not tell us that.

Premise 3: The intervention can play the same role *here* as it did *there*. The RCT evidence does not tell us that.

Premise 4: The enabling factors necessary for the intervention to play a positive causal role for Y are in place here, or we can get them. The RCT evidence does not tell us that.

Premise 5: The system (context) *here* is stable enough so that the intervention will have time to unfold and work. We know the main factors influencing this stability and we know how to maintain them. The RCT evidence does not tell us that.

Conclusion: Yes, the intervention will most likely work here. There are always unknown factors that might disable or hinder its workings; despite these we think it is worth implementing it. Or we may conclude that since the vital enablers are missing and they are too expensive to get, chances are that this intervention will not contribute positively to Y in our context.

This tentative argument structure can guide you to what kind of evidence you need to ascertain. As should be plain, the RCT evidence alone will not be enough.

5. *Conclusion*
I have in this paper addressed one aspect of evidence-based practice, namely the

fact that a lot more evidence is required in practice than is normally assumed by proponents and critics of EBP alike. The EBP literature, whether written by critics, adherents or researchers, focuses on RCT evidence as the kind of evidence on which practice should be based. Organizations such as CampbellCollaboration and McREL, which collect and vet evidence and produce meta-analyses, adhere to the EBP orthodoxy and the evidence hierarchy and view RCTs if not as the only admissible kind of evidence, then certainly as the preferable kind of evidence. Critics problematize this view point and argue that other kinds of evidence should count as well.

What none of them do, I have argued, is to address the question of what practitioners really need evidence *for.* If we assume that what practitioners really want to know is whether a proposed intervention will work for *them,* in their classroom, then it immediately transpires that RCT evidence is not enough. There are two reasons for this. The first is that RCT evidence only pertains to the first of the five premises I have suggested above. The second is that contrary to popular belief, RCT evidence does not show that a causal relation (X-Y) holds in general, it just shows that is holds for the study group from which the evidence emanates. In order to make a decision about whether we actually should implement the intervention in question *here*, we need to collect a good many local facts and put all our evidences together in an argument structure which allows us to make a sensible all-things-considered judgment. We must never lose sight of the fact that here denotes an already existing practice, a causal system, and that any output has many antecedent events. Changing a factor in the system or inserting a new one will bring changes to the entire system; changes which may affect our desired outcome in good or bad ways. RCT evidence may be highly trustworthy, but it does not even provide half the story. Putting all the different kinds of evidences in a structure will help us think systematically about what we need to know. Thus, EBP as a practical enterprise is indeed well served by setting all the necessary evidences in an argument.

I would like to end this paper with a remark about EBP itself: EBP is much more complicated that advocates and critics alike tend to think. It is essential to distinguish between finding and using causes, and it seems to me that using them to bring about desired results is much more complicated than finding them in the first place. EBP is thus no magical bullet for improving student achievements, but nor is it impossible. As a minimum it requires practitioners who can think for

themselves; the EBP orthodoxy is seriously misguided.

References

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

Arnesen, A., Ogden, T. & Sørlie, M. (2006). *Positiv atferd og støttende læringsmiljø i skolen [Positive behaviour and a supporting school environment].* Oslo: Universitetsforlaget.

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

Cartwright, N. (2009). How to do things with causes. *Proceedings and Addresses of the American Philosophical Association,* 83, 2, 5-22.

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

Hitchcock, C. (2007). Prevention, pre-emption, and the principle of sufficient reason. *Philosophical Review,* 116, 4, 495-532.

Kelly, T. (2008). Evidence. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/evidence/ Retrieved September 16, 2008.

Menzies, P. & Price, H. (1993). Causation as a Secondary Quality. *British Journal for the Philosophy of Science,* 44, 187-203.

Pawson, R. (2012). *Evidence-based policy. A realist perspective.* Los Angeles: Sage.

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

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

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

Woodward, J. (2008). Causation and manipulability. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy,* http://plato.stanford.edu/entries/causation–mani/ Retrieved March 20, 2012.

US Department of Education (2003). *Identifying and implementing educational practices supported by rigorous evidence: A User Friendly Guide.* Washington, DC: Coalition for Evidence-Based Policy. http://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf. Retrieved June 12, 2014.