THE CONCEPT OF EVIDENCE IN EVIDENCE-BASED PRACTICE

Tone Kvernbekk

Department of Educational Research
University of Oslo

Abstract. There exists a vast literature on evidence-based practice (EBP) in education. The debate branches out in several directions, for example, what EBP entails for the nature of educational practice, what it entails for the teaching profession, what counts as use and abuse of evidence, and what educational research could or should contribute to a *what works* kind of practice. In this essay Tone Kvernbekk focuses on the fate of the concept of evidence in the debate, observing that the concept seems, by and large, to be missing from the debate. She argues that educational debates about EBP stand to gain in nuance and depth from employing philosophical insights about evidence. Kvernbekk develops this claim by discussing different conceptions of evidence and by inquiring into three aspects of the evidentiary relation: the meaning of "based," underdetermination, and the relativization of evidence.

Introduction

There exists a vast literature on evidence-based practice (hereafter EBP), hardly surprising given the status of "evidence-based" as a buzzword in contemporary educational debates (and also in medicine and policymaking, among other areas). A cursory glance at the EBP literature reveals that in education, EBP is mainly discussed within a political context. Both British and American authors are responding largely to the demand for better research bases to inform both policy and practice, which has become known as the *what works* agenda.[1] While I am going to employ an epistemological perspective in my discussion of EBP, I would like to suggest very briefly that a rhetorical analysis of how the term "evidence" operates in discussions of EBP also would be of great interest. To be able to appeal to the existence of (empirical) evidence might give an aura of scientific support to views or policies that is misleading, perhaps even unfounded, depending on the quality of the evidence. Furthermore, in educational research conclusions tend to be contradicted in other studies; that is, in a good many cases there is both positive and negative evidence. The weighing of evidence is itself a complex process and decision makers (and others) may be very selective in their appeal to evidence in order to support or justify their views — proponents of different sides in virtually any debate can thus claim that the "evidence" supports their view. Negative evidence runs the risk of simply being ignored. As D.C. Phillips suggests, "evidence" may be used to justify decisions already made on other, more surreptitious grounds and to present a case as having a stronger foundation than it actually does and thus manipulate the public.[2]

1. See, for example, Helen Simons, "Evidence-Based Practice: Panacea or Over Promise?" *Research Papers in Education* 18, no. 4 (2003): 303–311.

2. D.C. Phillips, "Adding Complexity: Philosophical Perspectives on the Relationship Between Evidence and Policy," in *Evidence and Decision Making: 106th Yearbook of the NSSE*, ed. Pamela Moss (Oxford: Blackwell, 2007), 379.

Roughly, there seem to be two interrelated branches of EBP discussions: what educational research can and should contribute to a *what works* kind of practice, and what EBP entails for the nature of educational practice and the teaching profession. In the EBP conceptual landscape we find such concepts as relevance, effectiveness,[3] technicality, rationality, instrumentality, causality, randomized controlled trials (hereafter RCT), and positivism.[4] Against this cluster of concepts, critics pit another cluster of concepts consisting of practical judgment, professional experience, situatedness, appropriateness, ethical considerations, *phronesis*, and democracy — not infrequently with the tacit assumption that the two clusters are incompatible and that the latter is in danger of being completely replaced by the former.

Admittedly, the preceding characterization of the EBP debate is very brief and may not do justice to all aspects of the debate. But insofar as it captures the main aspects, one thing stands out: the concept of evidence is conspicuously absent. And that is what interests me in this essay: the fate of *evidence* in the EBP debate. It seems to me that EBP has taken on a number of connotations that have been shaped by the political context in which it was first discussed. Here I would like to strip these connotations away in order to get a better look at the concept of evidence itself. For if we look at the three words "*evidence-based practice*," they appear innocent enough. There is nothing in those words that makes it obvious or necessary that they entail a technical and instrumental practice that is solely preoccupied with measurable outcomes and ignores questions of the appropriateness of goals, as Gert Biesta has argued, for example.[5] There is nothing inherently positivist in the words *evidence-based practice*, nor do they by themselves preclude professional experience or the importance of situational appraisals by practitioners.

I shall turn to philosophy of science to get a grip on the concept of evidence. A cursory look at the philosophical debates about the nature and role of evidence reveals that in this conceptual landscape we find such concepts as evidence, hypotheses, beliefs, background, justification, support, good reasons,

3. David Hargreaves, "Teaching as a Research-Based Profession: Possibilities and Prospects," lecture (London: Teacher Training Agency, 1996); see also David Hargreaves, "In Defense of Research for Evidence-Based Teaching: A Rejoinder to Martyn Hammersley," *British Educational Research Journal* 23, no. 4 (1997): 405–419.

4. These are voiced by critics of EBP such as Gert Biesta, "Why 'What Works' Won't Work: Evidence-Based Practice and the Democratic Deficit in Educational Research," *Educational Theory* 57, no. 1 (2007): 1–22; John Elliott, "Making Evidence-Based Practice Educational," *British Educational Research Journal* 27, no. 5 (2001): 555–574; Martyn Hammersley, "Educational Research and Teaching: A Response to David Hargreaves' TTA Lecture," *British Educational Research Journal* 23, no. 2 (1997): 141–161; and Ian Sanderson, "Is It 'What Works' That Matters? Evaluation and Evidence-Based Policy-Making," *Research Papers in Education* 18, no. 4 (2003): 331–345.

5. Biesta, "Why 'What Works' Won't Work," 6–8.

TONE KVERNBEKK is Professor in the Department of Educational Research at the University of Oslo, P.O. Box 1092 Blindern, N-0317 Oslo, Norway; e-mail <Tone.kvernbekk@ped.uio.no>. Her primary areas of scholarship include narrative theory and the application of epistemology to various problems in educational theorizing.

confirmation, disconfirmation, truth, induction, and probability. Thus an initial glance at the two landscapes or debates would seem to indicate that there is no overlap between them, except for the concept of evidence itself. Nevertheless, I argue that educational debates about EBP stand to gain in nuance and accuracy from employing philosophical insights about evidence. First, I look at various conceptions of evidence and its functions. Second, I inquire into selected aspects of the evidentiary relation: the meaning of "based" in evidence-based practice, underdetermination, and the relativization of evidence. While these three issues, I submit, are highly pertinent to the EBP discussion, other equally interesting issues must be left untouched, most notably induction and probability.

Conceptions of Evidence

I begin this section by invoking David Gamson in support of my contention that the concept of evidence is lacking from the EBP debate. While he himself discusses *how* evidence historically has been used by educators to improve practice, he also gives an overview of other central questions in the EBP debate:

> The first focuses on the question of *what* has constituted "evidence" in the past and leads, in turn, to a second question about *who* educational leaders believed could legitimately conduct research, collect evidence, and take action on findings. A third issue relates to questions of *why* certain types of evidence were privileged at various points in our history.[6]

Gamson's questions of *who*, *why*, and *what constitutes evidence* nicely capture many of the issues and concerns discussed by both critics and advocates of EBP. But one *what* question is absent from Gamson's overview: the question of what evidence *is*. I have no hypothesis as to why this is so; maybe the concept is taken for granted. It is to the question of the nature of evidence that we now turn.

Peter Achinstein thinks that philosophical theories of evidence are generally ignored by scientists, that is, that scientists disputing whether or to what extent a putative piece of evidence supports a hypothesis do not turn to philosophy for help.[7] This sad state of affairs, he suggests, is largely due to two assumptions that philosophers tend to make about evidence. First, that evidence is a weak notion: very little evidence is needed to establish that something is the case. Second, that the evidentiary relation is a priori, not empirical. It is a logical, mathematical relation that is established by calculation, not by empirical investigation. Neither assumption fits scientific practice. Thomas Kelly also identifies a tension between philosophical and commonsense conceptions of evidence.[8] The commonsense conception mainly centers on physical objects indicating what happened — as, for example, when sticky fingers provide evidence of who raided the raspberry jar or ate the chocolate cake. This conception is not easily recognizable in historically

6. David Gamson, "Historical Perspectives on Democratic Decision Making in Education: Paradigms, Paradoxes, and Promises," in *Evidence and Decision Making*, ed. Moss, 15–45.

7. Peter Achinstein, *The Book of Evidence* (Oxford: Oxford University Press, 2001).

8. Thomas Kelly, "Evidence," in *The Stanford Encyclopedia of Philosophy (Fall 2008 edition)*, ed. Edward N. Zalta, http://plato.stanford.edu/entries/evidence.

prominent accounts of the nature of evidence. Empiricists, for instance, have seen evidence as consisting of sense data or protocol sentences: mental and linguistic entities, respectively, and most likely nothing that researchers will recognize from their practice or find useful. Philosophical and ordinary talk about evidence seem divided by a gulf, Kelly suggests.

The diagnosis is that philosophical theories of evidence have tended not to provide scientists with what they need. To remedy this state of affairs, Achinstein sets out to provide a robust conception of evidence that will be useful to researchers. The root meaning of evidence for Achinstein, as for Kelly, is *that which justifies belief*. Evidence is something that supports the truth of a hypothesis or indicates its falsity. However, as Larry Laudan points out, the story may be more complex than that. He distinguishes four different positions in which a theory (belief, claim) can be said to be supported by evidence.[9] The theory may

- be logically compatible with the evidence,
- logically entail the evidence,
- explain the evidence, or
- be empirically supported by the evidence.

Arguably none of these relations reduces to any of the others, according to Laudan. Of special interest for the topic of this essay is his argument that satisfaction of either the compatibility relation or the entailment relation fails to establish a relation of support between theory and evidence. Laudan's main target is the underdetermination thesis, which I shall come back to in a subsequent section. In my analysis here, I choose to follow Achinstein and focus on evidence in its main function, as something that supports a theory.

The function of support can in principle be performed by facts, experiences, and all sorts of data; there is nothing in the concept of evidence to privilege RCT and quantitative data at the expense of all other forms of data. Evidence confirms or disconfirms a hypothesis (theory, claim, belief). The key word to be remembered is *support*. It should be noted that this evidentiary relation largely assumes that the evidence is true, factual, accurate, or at least trustworthy — this seems to be taken for granted. Within this basic meaning of evidence, Achinstein identifies four different but related concepts of evidence, all of which, he says, are used in science. Let us look at the four types.

Achinstein's first type of evidence is called *epistemic situation evidence*: evidence that is relativized to an epistemic situation (ES). This kind of understanding allows us to ask whether researcher X was justified in believing hypothesis Y, given what else this researcher knew, what he or she did not know, what he or she could not know because it was only discovered later, and so on. It

9. Larry Laudan, *Beyond Positivism and Relativism: Theory, Method, and Evidence* (Boulder, Colorado: Westview Press, 1996).

requires, of course, that it is possible to reconstruct people's epistemic situation. It is also a fallibilist conception: you may be justified in believing a hypothesis even if it is false. It is important to note that ES-evidence is not subjective. Whether E is evidence that H or not does not depend on whether anyone actually believes this hypothesis. E would be evidence that H for anyone in such an epistemic situation.

Subjective evidence, on the other hand, Achinstein's second type of evidence, concerns how evidence is *regarded*. Subjective evidence is relativized to the person or group, and the crucial point is how the putative evidence is regarded by the group. The following elements are involved: The group (or person) believes that E is evidence for H; it believes that H is true, and the group's reason for believing that H is true is that E is true. However, the subjective conception of evidence does not require that E is in fact true, only that the group believes it. Furthermore, the group does not have to be justified in its belief that H; the crucial point is what the group itself believes.

Achinstein's third conception of evidence is termed *veridical*. This is a thoroughly epistemic conception: If E is evidence that H, then E provides a good reason to believe H. It is an objective notion of evidence; whether E is a good reason does not depend on anybody actually believing it. The good reasons may be empirically incomplete, but if E is a good reason to believe H, then H is true. As one can see, this is a very strong conception of evidence; it requires that both E and H are true. But is such evidence attainable in scientific practice? It certainly seems to be what some advocates of EBP have in mind, even if their wording is different. In his now-famous lecture sponsored by the Teacher Training Agency, David Hargreaves argues that educational research should be able to provide decisive and conclusive evidence that teaching strategy X rather than strategy Y yields significant improvement in outcome.[10] Clearly he must assume that this evidence is both veridical and objective; it does not depend on what anybody believes about it. But decisive and conclusive are even stronger than veridical. Achinstein takes great care to distinguish between veridical and conclusive evidence, intimating that conclusive is *too* strong. If E is conclusive, then not only is H true, but E establishes it with certainty. Perhaps conclusive evidence can be seen as a limiting case of veridical evidence.

Veridical evidence, Achinstein suggests, is the kind of evidence that scientists seek. Only providing justification of belief for those in certain epistemic situations is not enough, since the belief might turn out to be false. Scientists want their hypotheses to be true both generally and objectively, and veridical evidence obviously transcends the concrete situations in which it was gathered. Suppose we have a teacher emphatically denying the usefulness of a research-based strategy, claiming that she has experiential evidence that another strategy works much better. In such cases Achinstein's vocabulary provides an elegant explanation: If E is the teacher's subjective evidence that H, then she tends to believe that E is veridical evidence that H.

10. Hargreaves, "Teaching as a Research-Based Profession."

However, veridical evidence is too strong a notion. Achinstein's fourth type of evidence, which he eventually settles on as the basic form, is called *potential*. It is objective — that is, it requires E to be true — and it is connected to good reasons, but in a weaker sense than veridical evidence. This is a fallibilist conception in that it allows H to be false even when there exists evidence to support it.

So what does all this tell us about the nature and function of evidence? It tells us that evidence is something that has a bearing on the truth-value of a hypothesis (theory, belief); in other words, it is something that supports or confirms the hypothesis, justifies our belief in it. Or the opposite: evidence can also disconfirm a hypothesis. It is important to note that confirmation and justification do not entail truth, as Achinstein makes clear. Verification is the limiting case of confirmation; it conclusively establishes the hypothesis as true. Falsification is the limiting case of disconfirmation.[11] What we end up with if we accept Achinstein's view is an epistemic conception of evidence that explains why attention to evidence is important for people who are concerned with building true pictures of the world.

The nature of the relation between a piece of evidence and that which it is evidence for has been a matter of much philosophical discussion, hardly any of which is reflected in the educational EBP debates. So how does evidence bear on hypotheses?

The positive relevance view is a probabilistic view of evidence, claiming that E is evidence for H only if E increases the probability of H (of H's being true).[12] Another way of putting this is to say that E provides inductive support for and confirms H; new pieces of positive evidence add to the probability of H. At the outset, this allows for a hypothesis to be more or less confirmed — a consideration that should be highly relevant in EBP contexts, given the complexity of educational phenomena. Confirmation is not an all or nothing affair, but a matter of degree; and this degree increases with more evidence. This viewpoint seems intuitively appealing and might work well for many purposes. Nevertheless, a good many philosophers have argued that the positive relevance view fails as a principled view of the evidentiary relation, specifically, Achinstein, Edward Erwin and Harvey Siegel, and Clark Glymour.[13] They basically agree in their criticism: there are cases in which the increase in probability simply is too small for it to be reasonable to speak of confirmation. It is evidence, but too weak to confirm the hypothesis. Thus positive relevance fails to provide a sufficient condition for evidence. Any infinitesimal increase in probability does not imply an increase in the strength of evidence; there is no automatic increase in confirmation.

11. Kelly, "Evidence."

12. Wesley Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, New Jersey: Princeton University Press, 1984).

13. See, for example, Achinstein, *The Book of Evidence*; Edward Erwin and Harvey Siegel, "Is Confirmation Differential?" *British Journal of Philosophy of Science* 40, no. 1 (1989), 105–119; and Clark Glymour, *Theory and Evidence* (Princeton, New Jersey: Princeton University Press, 1984).

Interestingly, these critics also largely agree about the remedy, even if they have different words for it. Achinstein speaks of thresholds and Erwin and Siegel speak of differentialness. Achinstein makes high probability a necessary but not sufficient condition for a robust conception of evidence. By "high" he means "larger than one-half," a threshold to which evidence should raise the probability of H if it is to provide a good reason to believe H and thus make it at all interesting for science — or for professional uses, one might add. Just how one judges probability in a given case, Achinstein does not say; that is left to the researchers. What is important is that any little piece of evidence is not enough to provide reasons to believe in a hypothesis. Erwin and Siegel argue that the one-half standard is not arbitrary at all; its rationale is provided by the requirement of differentialness. The effect is the same: if a hypothesis enjoys a probability larger than one-half, then no rival hypothesis that is incompatible with it can have a probability higher than that, and so we have good reason to believe it. I shall come back to this point; it is important in order to understand both the allure of RCTs and the underdetermination thesis.

But first we need to take a brief look at another standard theory of (objective) evidence: the hypothetico-deductive (H-D) account. On this view, for E to be evidence that H it suffices for E to be deductively derivable from H (or from the conjunction of hypothesis and background). That is to say, you confirm a hypothesis by deriving a prediction that is subsequently discovered to be true by observation. This view, I suspect, sounds familiar to most researchers. But the H-D account, most writers agree, yields a very weak notion of evidence indeed. This is so because the same fact may be derived from a number of different, even mutually incompatible hypotheses. This fact would then be evidence for all of these rival hypotheses — a result that is counterintuitive. As we shall see, the H-D account is closely connected with the underdetermination thesis, a thesis that is appealing to many and that may cause serious problems for EBP.

The Evidentiary Relation

The positive relevance view, the threshold/differentialness view, and the H-D account all concern how evidence is related to that for which it is evidence. If we further unpack the evidentiary relation, we find issues of great interest — confirmation, induction, underdetermination, probability, and the like — all of which in one way or another pertain to EBP but that are by and large left untreated by educational researchers and policymakers discussing EBP.

It is important to note that the philosophical discussion so far centers on the relation of evidence to single, isolated hypotheses. It is by no means self-evident that the relation between evidence and practice can be modeled on the relation between evidence and isolated hypotheses. "Practice" is, I surmise, an entity that does not admit of exactly the same kind of relation to evidence as isolated hypotheses do, especially if one understands practice to be a complex social activity with long traditions and its own aims and standards. As both advocates and critics of EBP point out, what is at stake in EBP is the *effectiveness* of practice, not the truth of hypotheses. Effectiveness and truth are not the same

thing. Effectiveness concerns the attainment of goals, nowadays often evaluated by how measured outcomes compare to prespecified goals. Keeping this caveat in mind, it is still legitimate, of course, to speak of a certain practice being supported or justified by evidence. Indeed, practice is something that generally stands in need of justification, but given its complex mix of social, normative, and empirical elements, it seems reasonable to suggest that several types of evidence and reasons are needed for such justification. While empirical evidence for practice may concern its effectiveness, practice also needs to be justified by moral, social, and educational reasons, among others. But while the effectiveness of practice and the truth of hypotheses are different things, the function of evidence is the same: support. For example, E may support the effectiveness of teaching method X over that of method Y. On Achinstein's potential evidence concept, this would amount to our having good reason to believe that method X is better than method Y, acknowledging the possibility that this might be wrong. If the evidence suggests that both methods have equally high probability, then other factors enter into the evaluation, such as cost, social desirability, and so on.

On the Meaning of "Based" in Evidence-Based Practice

I have argued that one basic meaning of "evidence" involves confirmation and disconfirmation of hypotheses (beliefs, claims), summed up in the keyword *support*. But this does not seem to be what most advocates and critics of EBP alike take evidence to be (nor, for that matter, do they seem take it to involve any of the other three relations between theory and evidence mentioned by Laudan). Strictly speaking, the phrase "evidence-*based* practice" may be a little misleading, in that it suggests that practice is based on evidence. But the practice in EBP is not really based on evidence; it must rather be understood to be based on the hypothesis or theory for which the evidence is evidence. This may seem like an unnecessarily subtle distinction, and the two may indeed sometimes be hard to separate when educational practice is concerned. But the distinction between evidence and theory is important to maintain in EBP as well as in other contexts. Sometimes the evidence looks very different from the claim it supports. For example, in a study of the differential effects of two reading intervention programs, the evidence consists of tables, figures, numbers, and descriptions detailing various test scores. The theory tells us that a method focusing on oral language training is effective for teaching poor readers vocabulary and grammatical skills, whereas a method focusing on phonology with reading is more effective concerning literacy.[14] Or, to take a more mundane example: the evidence consists of a blood spatter pattern, a credit card receipt from a gas station, and two witness statements; but the theory is that the butler did it. Since this is a fallibilistic conception of evidence, the butler may prove innocent after all.

14. Claudine Bowyer-Crane et al., "Improving Early Language and Literacy Skills: Differential Effects of an Oral Language versus a Phonology with Reading Intervention," *Journal of Child Psychology and Psychiatry* 49, no. 4 (2008): 422–432.

Bronwyn Davies is an ardent critic of EBP and one of very few who explicitly discusses what "based" might mean.[15] Her view of this is illuminating. She thinks of "based" in terms of a base or foundation consisting of facts and (quantitative) data. I think she voices what a good many of those writing about EBP take "based" to mean but do not say explicitly. There is a tendency to understand it as *derivation*, in the sense that practice could and should be derived from a foundation of evidence. For her part, Davies focuses on which evidence should make up the base and on what Gamson calls the *who* question, which is closely connected to issues of power: who should select the evidence and thus decide what is relevant and what is not. Biesta argues that the most extreme advocates of EBP understand "based" in terms of derivation. He describes them as "those who think that research will be able to give us 'the truth,' that 'the truth' can be translated into rules for action, and that the only thing practitioners need to do is to follow these rules without any further reflection on or consideration of the concrete situation they are in."[16] Needless to say, Biesta strongly disagrees with such a view. Some observations are in order. First, empirical research results by themselves do not provide practitioners with specific recipes for action; I agree with Biesta here. Second, a possible derivation of practice or hypothesis from evidence is not an instance of the H-D method, although it may look like it. On that account, evidence is derived from a hypothesis, not vice versa. Third, and most importantly for my purposes in this essay, the derivation view mistakes the function of evidence as I described it in the previous section. A hypothesis, or a teaching strategy, is not *based* on evidence; instead, it is *supported* by it.

Admittedly, this is tricky ground. We often speak of theories as generated on the basis of observation and data. Phillips has a rather lengthy discussion of the meaning of "based."[17] He explicitly states that the relation of a hypothesis to its putative evidence is one of support, not foundation. Facts (evidence) cannot be regarded as a base from which theory, policy, or practice can be inferred. Phillips uses Sir Isaac Newton's theory of gravitation as an example: his theory is not *based* on data; the theory cannot be derived either deductively or inductively from the data that were available to him. Rather, Phillips suggests, the data were a source of puzzlement, and Newton could plausibly have asked himself how the data could be explained, what unobservable mechanism could possibly be at work that would produce such data. Phillips concedes that we in some sense might say that data generate hypotheses, but in the indirect manner described.

Summing up, the word "based" in evidence-based practice has led several writers to think of evidence as a foundation, as knowledge in itself or as a more or less well-circumscribed collection of facts from which teaching methods or rules of action can be inferred. This is not how philosophical theories portray the function

15. Bronwyn Davies, "Death to Critique and Dissent? The Policies and Practices of New Managerialism and of 'Evidence-Based Practice'," *Gender and Education* 15, no. 1 (2003): 91–103.

16. Biesta, "Why 'What Works' Won't Work," 11.

17. Phillips, "Adding Complexity," 381–387.

of evidence, and many misunderstandings about EBP — conceivably held by both advocates and critics — might have been avoided if philosophical insights had been employed. The evidentiary relation is not one of foundation; it is one of support.

The Thesis of Underdetermination

To the best of my knowledge, neither advocates nor critics of EBP discuss underdetermination, with the exception of Phillips. Underdetermination centrally concerns the relation of evidence and theory, and would provide good ammunition for the EBP critics. The thesis is intended to convey the notion that adoption of a theory cannot be based on considerations of evidence or facts alone; facts underdetermine theory. Underdetermination in effect means that there is a gap between evidence and theory, policy, or practice that must be filled with other data, values, various assumptions, and contextual information. This thesis should thus fit the critics of EBP well, since they all argue precisely that practice cannot be generated on the basis of evidence alone, or that evidence alone cannot support a theory. Underdetermination is therefore a serious threat to EBP because it undermines the idea of evidential support.

Phillips does not write in detail about underdetermination, but his main point is that theories transcend the evidence that is available. This means that the same data or evidence might be compatible with more than one theory, even if the theories themselves are incompatible. In Newton's case the data are also compatible with a rival theory, namely Albert Einstein's theory of relativity. The principle of underdetermination has an undeniable appeal to it, and Phillips takes its validity for granted. Indeed, the thesis of underdetermination seems to be a truism in many scientific studies. But things may not be so simple. According to John Norton, the thesis maintains that all evidence necessarily underdetermines any scientific theory, a view that is easily generalized to any practice or policy.[18] There is thus an *assured* possibility of rival theories that are at least as well confirmed as the original theory. It is important to note that the thesis implies that rival theories are *equally well* confirmed by the evidence, such that choice between them must be due to other factors, such as values, subjective preferences, or contextual information. Furthermore, it is not the case that the evidence at hand just happens to leave a theory underdetermined; the thesis says that underdetermination will persist in all cases no matter how much evidence you collect. Norton states that this is simply what the thesis of underdetermination says, that this is what it means in all cases. It seems to me, however, that one can see this as a strong form of underdetermination and that there also might be a weaker form. The weak thesis of underdetermination would say, then, that while the evidence is compatible with several theories, it does not mean that the rival theories are equally well confirmed. It is not clear to me whether Phillips would accept the strong form of underdetermination and regard all possible rival theories

18. John D. Norton, "Must Evidence Underdetermine Theory?" in *The Challenge of the Social and the Pressure of Practice,* ed. Martin Carrier, Don Howard, and Janet Kourany (Pittsburgh, Pennsylvania: University of Pittsburgh Press, 2008), 17–44.

as equally well supported and hence equally believable. His general Popperian instincts make all talk of confirmation difficult; he describes the role of facts as being a source of or contextual background for problems, not of confirming hypotheses.

Two of the best-known proponents of the underdetermination thesis are Pierre Duhem and Willard Quine, both of whom subscribed to holistic versions of it.[19] Duhem's point is that a hypothesis cannot be tested in isolation, since the conjunction of hypothesis and background is necessary to derive testable observations. If the observation fails to obtain, we are free to accommodate theory and background in many ways. And if this is the case, then we have many different systems of hypotheses that accommodate the observation. Quine held a strong version of underdetermination as a result of his famous rejection of the two empiricist dogmas: the analytic/synthetic distinction and the notion that meaningful statements are reducible to statements of immediate experience. According to Quine,

> The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion ... is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.[20]

Science, Quine argued, is a fabric that impinges on experience only on the edges. Any conflict with experience occasions reevaluations in the interior of the fabric, and the reevaluation of some statements entails reevaluation of others because of their logical interconnections. He went on to claim,

> But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.[21]

What we see in both Duhem and Quine is an idea of a distance between experience or evidence and hypothesis, such that the evidence is unable to fix the hypothesis uniquely. Underdetermination, in Quine's view, permeates our entire conceptual systems as a matter of fact. A statement in the field that is threatened by some piece of recalcitrant evidence can always be rendered immune by adjusting some other statements. This leads to Quine's extraordinary conclusion: "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system."[22] Laudan suggests that this must be taken to mean that "it is rational to hold on to any theory whatever in the face

19. See, for example, Pierre Duhem, *The Aim and Structure of Physical Theory* (Princeton, New Jersey: Princeton University Press, 1954); and Willard Van Orman Quine, *From a Logical Point of View: Nine Logico-Philosophical Essays*, 2nd rev. ed. (Cambridge, Massachusetts: Harvard University Press, 1980).

20. Quine, *From a Logical Point of View*, 41.

21. Ibid., 42–43.

22. Ibid., 43.

of any evidence whatever''[23] — negative evidence never disconfirms a theory. If Quinean underdetermination is accepted, then the basic idea behind EBP — choose the theory that is best supported by the evidence — is thoroughly undermined. So on this thesis, if the effectiveness of teaching method A is confirmed by evidence, then the same evidence is compatible with some other method and supports or confirms it equally well. The teacher's choice of method is ultimately made not on evidence, but on other contextual considerations and perhaps personal preferences.

Norton argues that underdetermination clearly makes a claim about confirmatory relations, but it relies on a flawed account, namely the H-D model of confirmation. As already noted, the H-D model is troubled by an excessive permissiveness. It is precisely this permissiveness that arguments for underdetermination seek to exploit, and Norton points out that the H-D model is routinely invoked in accounts of underdetermination. If the same fact can be derived from different systems of hypotheses or fields, then the fact will be evidence for both and thus confirm both, even if the systems are incompatible. Norton, for his part, concludes that the thesis of underdetermination enjoys a status it by no means deserves, premised as it is on a flawed conception of confirmation. If underdetermination were true, any far-fetched story that can be constructed to fit the facts is just as good as a scientific theory that fits the same facts, and this will not do in scientific practice. Laudan maintains that even if two theories enjoy the same set of known positive instances, it does not follow that they should be regarded as equally well supported or confirmed by those instances. To say that rival theories enjoy the same empirical support, he argues, requires more than just saying that the theories are compatible with, or capable of entailing, the same evidence. Thus, one of his main criticisms of Quine is that Quine failed to distinguish between the four different relations of theory and evidence cited earlier.

Achinstein's view of evidence and confirmation runs counter to the underdetermination thesis in two ways.[24] First, Achinstein allows for the confirmation of isolated hypotheses while Duhem and Quine both held holistic views.[25] For Duhem, it is the conjunction of hypothesis and background assumptions that is confirmed. For Quine, as we have seen, the field of science meets experience as a whole. Although, to be fair, it should be noted that Quine later relaxed his holistic views. Second, Achinstein requires that the evidence confer a high probability on the hypothesis. If a hypothesis enjoys a probability larger than one-half, then no rival incompatible theory can be equally well confirmed by the same evidence. The high probability requirement is not compatible with the strong version of the underdetermination thesis. Achinstein's requirement echoes the views of Erwin

23. Laudan, *Beyond Positivism and Relativism,* 34.

24. Achinstein, *The Book of Evidence.*

25. He regretted his ''needlessly strong statement of holism'' and argued instead for a moderate holism that invokes ''not the whole of science but chunks of it.'' Willard Van Orman Quine, ''Two Dogmas in Retrospect,'' in *Quintessence: Basic Readings from the Philosophy of W.V. Quine,* ed. R.F. Gibson (Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2004), 57.

and Siegel.[26] Confirmation, they argue, is *differential*: E confirms H if E gives at least some reason for taking H to be true and, at the same time, does *not* afford equal or better reason for believing a rival hypothesis. In some sense, Erwin and Siegel argue, confirmatory experiments are crucial experiments; that is, they differentiate between a hypothesis and its plausible rivals. This does not imply that such experiments always yield conclusive evidence; evidence can be relatively weak and still be evidence. It would seem that a good many quantitative studies are designed on this logic of differentialness, since they often provide comparisons of two or three rival hypotheses with the explicit intention of differentiating between them. For example, take the teaching of reading.[27] I previously referred to a study by Claudine Bowyer-Crane and colleagues focused on the differential effects of two reading intervention programs; in another recent study, Peter Hatcher and fellow researchers examined the effectiveness of small-group interventions for students demonstrating reading delays. Both of these studies compare the effectiveness of reading interventions. Hatcher's study is an RCT with a twenty-week intervention group and a ten-week intervention group. The results reported indicate that the intervention is effective for 75 percent of children who are showing reading delays at the end of their first year in school; the other 25 percent did not respond. In Bowyer-Crane's study, which also uses an RCT approach, one type of intervention gave better literacy and phonological measures, whereas the other intervention had better results in terms of vocabulary and grammatical skills. For a practitioner wondering what to do about grammatical skills, this RCT gives the edge to one of the interventions. None of the interventions, it should be noted, are "perfect." A particular strategy may work for many, but not for all.

Arguably, this logic also lies behind policymakers' apparent love for RCTs, since they precisely offer an opportunity to rule out some hypotheses while providing differential confirmation for others. We see that both Achinstein and Erwin and Siegel assume that evidence can play the role of neutral arbiter among rival theories. Surely RCTs proceed on the same assumption. If we have a situation where two hypotheses are equally well confirmed — that is, have the same probability — the rational response is to suspend belief.

To sum up, it seems to me that critics of EBP would have gotten much mileage out of the thesis of underdetermination had they employed it. It undermines the view often attributed to EBP that, to any given set of evidence, there belongs one unique teaching strategy or set of rules for action. In contrast, the thesis declares that evidence underdetermines theory (and, by extension, teaching strategy) such that the evidence confirms more than one theory (strategy). If, as Norton argues, the thesis holds that all theories compatible with this evidence are equally well confirmed, then choice of a theory must come down to contextual factors,

26. Erwin and Siegel, "Is Confirmation Differential?"

27. See, for example, Bowyer-Crane et al., "Improving Early Language and Literacy Skills"; and Peter Hatcher et al., "Efficacy of Small Group Reading Intervention for Readers with Reading-Delay: A Randomized Controlled Trial," *Journal of Child Psychology and Psychiatry* 48, no. 8 (2006): 820–827.

ethical considerations, or personal preferences. While this strong interpretation is problematic, I have suggested that a weaker interpretation of the thesis might be feasible, such that not all the theories are viewed as equally reasonable or believable despite being compatible with the evidence.

RELATIVIZATION OF EVIDENCE

As we have seen, two of Achinstein's four conceptions of evidence, ES-evidence and subjective evidence, imply a relativization of evidence to background (background broadly conceived to incorporate professional contexts and group or individual beliefs). While Achinstein concedes that all four conceptions are employed in science, he thinks that scientists seek veridical evidence whereas he himself settles on potential evidence as the basic form. Both are general, objective, and nonrelativized; both enjoy high probability; and both provide good reasons for belief in a hypothesis.

EBP critics would find it very difficult to accept Achinstein's veridical and potential types of evidence, I think, on the grounds that practitioners need context-bound knowledge, not generalized context-free knowledge. I shall make two observations here. The first is that proponents and critics alike might be guilty of a conflation here: evidence is conflated with the theory (knowledge) that the evidence is being used to support. It seems to me that in many cases, the production of evidence is seen as tantamount to the production of knowledge that is (thought to be) directly usable in practice. RCTs will provide general, objective *evidence* that may be taken for general, objective *knowledge*. As argued earlier, this mistakes the nature of evidence. Evidence is that which speaks to the truth-value of beliefs, theories, and claims, and an RCT is only one of several ways of obtaining evidence.

My second observation concerns practice and alludes to a long-standing debate in education, namely what kind of knowledge practitioners need. Practice deals in particulars and knowledge should therefore be relativized to different contexts. For example, John Elliott argues that since teacher judgments (teachers being the users of the results of educational research) are deeply context-dependent in nature, educational research should take the form of case studies rather than RCTs.[28] Ian Sanderson argues that what matters is not what works in general, but what is appropriate in particular circumstances, and this is the object of contextual judgment.[29] EBP deals in generalized knowledge and would thus, according to these arguments, simply be placed on the sidelines. Elliott and Sanderson are concerned with the nature of teaching as well as with forms of educational research in their defense of practitioner knowledge as contextual. Biesta takes a different angle and voices a thoroughgoing skepticism about general knowledge, arguing that we cannot in principle know what *works* generally, only what *worked* in a specific situation in the past.[30] It is not clear how this relates to *evidence*; that

28. Elliott, "Making Evidence-Based Practice Educational."

29. Sanderson, "Is It 'What Works' That Matters?"

30. Biesta, "Why 'What Works' Won't Work," 16.

is, we do not know whether Biesta conflates evidence and knowledge or whether he thinks we cannot have generalized, objective evidence of what worked in the past. Assuming Biesta speaks about evidence, which, if any, of Achinstein's four types of evidence is at play in his analysis? Clearly not veridical or potential since both allow us to have generalized, objective evidence of what *works*. Even ES-evidence is objective according to Achinstein's account, despite being relativized to an epistemic situation, in that it does not require anybody to actually be in that situation. I do not know if objectivity of evidence would be objectionable to Biesta; if it is, then we are left with subjective evidence — the question in this case perhaps becomes what the group of professionals regards as evidence. It is not clear to me why practitioners should restrict themselves to context-bound knowledge only and not avail themselves of general knowledge. It is understandable if one thinks that EBP entails that practitioners are reduced to following general rules of action, derived from equally general evidence. But, as I have argued, there is nothing in the words *evidence-based practice* themselves to suggest that this is the case. Advocates of EBP, such as Robert Herbert and colleagues, make it clear that the use of general knowledge implies judicious adaptation of this knowledge to the concrete circumstances.[31] The professional judgments of practitioners are therefore as necessary as ever in order to make the best use of their knowledge in serving the good of their students, patients, or clients. And as Achinstein points out, people tend to see their own subjective evidence not as subjective and fallible, but as veridical — as general, too, we might add, given that practitioners surely think that their experience-based beliefs are usable in similar situations in the future.

It should be noted that quite a few critics of EBP acknowledge that the EBP movement shares some of their concerns. As David Bridges and Michael Watts observe, "[EBP] is calling for practice (and policy) to be based on *evidence* as opposed perhaps to whim, prejudice or embedded custom."[32] Setting aside the troublesome meaning of "based on evidence," I wholeheartedly agree with the sentiment. But what happens when, say, a well-confirmed educational strategy for teaching children to read meets its intended public? How likely is it that evidence will replace experience-based beliefs and embedded custom? This is an issue that David Hargreaves also touches upon: how to make research results trustworthy for practitioners.[33] The problem may be bigger than proponents of EBP imagine, including Hargreaves and Herbert. A well-confirmed hypothesis is one that we have good reason to believe. Still, people may not find it believable, especially if it runs counter to preexisting values, commitments, and experience-based beliefs. As Phillips argues, believability is a function of coherence with our beliefs and values,

31. Robert Herbert et al., "Evidence-Based Practice — Imperfect but Necessary," *Physiotherapy Theory and Practice* 17, no. 3 (2001): 201–211.

32. David Bridges and Michael Watts, "Educational Research and Policy: Epistemological Considerations," *Journal of Philosophy of Education* 47, supplement 1 (2008): 44.

33. Hargreaves, "In Defense of Research for Evidence-Based Teaching."

not of rigorously conducted research, even RCTs.[34] That is to say, objectively speaking, belief in a strategy or teaching method may be highly reasonable, but we still might have no confidence in that strategy or method. This touches upon the political uses of "evidence" briefly mentioned in the introduction; decision makers (and others) may for ideological reasons choose not to believe the evidence (or they may select contradictory evidence), on the grounds that it does not fit their beliefs. Or the grounds for their rejection of the evidence may be more worldly, for example, if they ignore evidence because acting on it is judged to be too expensive.

Many researchers, philosophers, and proponents of EBP alike ascribe to evidence the role of neutral arbiter, a court of appeals to generate agreement. But even fully informed and rational researchers disagree, as the history of science shows. It is not unreasonable to generalize this to incorporate professional practitioners; they can also be expected to disagree about what the more effective strategy is and how one knows it. Is it the case, then, that the bearing of a given body of evidence on a given hypothesis or teaching strategy is a highly relativized affair? What, Thomas Kelly asks, must evidence look like if it is going to generate intersubjective agreement among inquirers or practitioners?[35] What constraints must we place on the sorts of things that are eligible to count as evidence in such cases? This is not exactly the same question that educationalists writing on EBP ask when they consider what counts as evidence (a question that is posed frequently in the EBP literature). Usually, they are concerned with widening the circle of what is considered as evidence from exclusive reliance upon RCT findings to include data from other forms of inquiry, as well as nonempirical considerations. Kelly asks what properties evidence must possess to play the role of neutral arbiter. Most importantly, he suggests, such evidence must be *public* — that is, it is available to, and can be grasped and appraised by, multiple individuals. And to be sure, both philosophers and researchers have championed such a view (for example, Carl Hempel and Karl Popper).[36] On the other hand, some philosophers have taken evidence to consist of sense data (for example, Bertrand Russell), and sense data are private, accessible only to the subject in question.[37] Such evidence cannot play the role of court of appeals, but it may still play a substantial part in decisions about teaching strategies — to reiterate, we tend to treat our own experiences and observations as veridical evidence.

In the work of Denis Phillips we find another way of relativizing evidence to background.[38] In the absence of a hypothesis, Phillips says, facts or pieces of data

34. D.C. Phillips, "A Guide for the Perplexed: Scientific Educational Research, Methodolatry, and the Gold versus Platinum Standards," *Educational Research Review* 1, no. 1 (2006): 15–26.

35. Kelly, "Evidence."

36. See, for example, Carl Hempel, *Fundamentals of Concept Formation in Empirical Sciences* (Chicago: University of Chicago Press, 1952); and Karl Popper, *The Logic of Scientific Discovery* (New York: Basic Books, 1959).

37. Bertrand Russell, *The Problems of Philosophy* (1912; repr. Oxford: Oxford University Press, 1997).

38. Phillips, "Adding Complexity," 386–387.

are *not* evidence. Depending on the formulation of a hypothesis, facts *become* evidence. In other words, evidence is made, not found. Just as Sherlock Holmes puts facts together into a coherent story about what took place when the crime was perpetrated (Phillips's own example), researchers put facts together into arguments, thus turning them into evidence. This view naturally goes well with the thesis of underdetermination. There are several possibilities here, Phillips argues. The same fact may be part of different arguments that reach divergent conclusions; arguments proceeding from different starting points and using different facts as evidence may arrive at the same conclusion; and the same fact as starting point may yield different endpoints. But should arguments utilizing the same facts as evidence but reaching divergent conclusions be viewed as equally good, as the thesis of underdetermination (in its strong version) would have it? Phillips does not say. What he explicitly states is that there is no rule of argumentation that constrains the types of evidence that can be used — any attempt at legislating RCT as the only evidence allowed is entirely misinformed about scientific history and practice, according to Phillips. While I agree with this latter point concerning RCTs, I would like to briefly problematize the notion that evidence is made.

I do not think that Phillips intends his views to endorse a subjective notion of evidence; that would go against his Popperian instincts and his emphasis on the importance of objectivity and fallibilism. Still, people may make different evidence of the same facts by constructing different arguments, by relativizing facts to argument (background). So if this "background" is not to be understood as subjective, given Phillips's general views, then what is it? I am not sure. Maybe Phillips's conception of evidence is akin to Achinstein's ES-evidence? It surely looks like evidence is relativized to a specific epistemic situation, namely the other facts and propositions that are known to me or to anybody in that situation. The problem with this is that for Achinstein, even ES-evidence is objective; it does not really require that anybody has those beliefs. I do not think that Achinstein would agree with Phillips that evidence is made. Admittedly, he does not discuss this problem directly, but he does discuss a problem that has a bearing on Phillips's views: whether something more than the truth of E is needed for E to be potential and veridical evidence for H. Must E also be *known* to be true? Clearly Phillips requires that E must be known. If nobody knows that E is true, then nobody knows that E is evidence that H, or that E could be put into a story or an argument. On Phillips's view, an unknown E would be uninteresting. But such an argument cuts no ice, Achinstein says. For if E was subsequently discovered, it would be of great interest. It would then be *recognized* as evidence. There is a distinction here between E's *being* evidence for H and our *recognizing* it as such. Achinstein's objective notion of evidence does not require that anybody actually believes it. E may confirm a hypothesis without anybody being aware of it, since being evidence for a hypothesis means that it supports the hypothesis. I think Achinstein would argue that facts are not *made* to support a hypothesis, but that they can be recognized as supportive (or unsupportive). And of course — and this is an important point — it is possible to believe falsely that E supports H.

In sum, relativization of evidence is a topic with many ramifications. A good many EBP critics argue that whatever evidence and knowledge practitioners need and use, it is context-bound and not general. It is not obvious what a "context" is taken to be: a classroom, a school, a community, or a country. Potential and veridical evidence both fall outside such a view. It might be that ES-evidence can accommodate such context-bound evidence, if local, contextual evidence is viewed as evidence for anybody who finds him- or herself in that context, whether the individual is aware of it or not. I have suggested that I believe that experience-based beliefs are also general and thought to be usable in more than one context. I certainly do not think that practitioners would happily settle for using subjective evidence to justify their choice of teaching methods; rather, we tend to treat our own experience as veridical evidence, justifiably or not.

Conclusion

In this essay, I have examined the concept of evidence in scholarship on EBP. While there is a substantial philosophical literature about evidence, it seems to be an under-analyzed concept in the educational context. I have inquired into three aspects of the evidentiary relation between evidence and that for which it is evidence: the meaning of "based" in evidence-based practice, underdetermination, and the relativization of evidence to background.

Evidence speaks to the truth-value of claims. In educational EBP contexts, it speaks to the degree of effectiveness of teaching strategies (or to the truth-value of our claims about such effectiveness). While it must be granted that evidence may have different, albeit interrelated, functions in relation to theory, its main function is that of support: confirmation or disconfirmation. Keeping that in mind, certain misunderstandings concerning the term "based" can be cleared up, and we can get past the discussion of whether we should be able to derive rules of action from the evidence. Facts, data, propositions, narratives, and the like can all constitute evidence, so attempts at legislating RCT data as the only legitimate form of evidence fail. What makes something evidence is that it stands in a certain relation to a hypothesis, namely confirmation or disconfirmation. EBP suggests that we choose a theory that is well supported, or a teaching method that is well supported. It means that we must have some idea of what "well supported" means, and to this end I have drawn on Achinstein and Erwin and Siegel. Admittedly, making such determinations is no simple task. But it is important, since any little scrap of evidence will not do to justify choice of theory or strategy.

In some respects this essay has taken the form of a defense of EBP. It is not my intention to suggest that none of the critics' concerns are legitimate. I certainly agree with the criticism that if education were focused exclusively on the goal of effectiveness, practice would be instrumentalized to an alarming degree. But I also think that my inquiry into the concept of evidence has indicated that much of what has been attributed to EBP and the idea of evidence is misguided. Philosophy has a contribution to make to the continued debates about EBP, for adherents and critics alike.