WHAT IS EVIDENCE?

Gary Thomas

Practice in Education. Maidenhead: Open University Press.)

Evidence in education and elsewhere

No one denies the importance of evidence in shaping and enhancing practice. At issue is
not the significance of evidence but its nature – and its value contingent on that nature. At
issue are the potency and value ascribed to certain forms of evidence in supporting
propositions that arise in educational practice. Many kinds of evidence are available to
practitioners in support of ideas and propositions that arise as part of their work: from
observation, from documents, from the word of others, from reason or reflection, from
research of one kind or another. It is the respect and worth credited to these that I shall
explore here as these issues of respect and worth seem to me to be at the core of what is
meant by ‘quality’ in educational research. I shall hinge my exploration around notions of
evidence in various different spheres of inquiry.

***

Mary was scanning the slope when suddenly she saw a tooth projecting from it,
just a speck of gray fossilized enamel. She looked once more and then shouted
for me to come.

Together we slowly cleared a little of the rock face with dental picks, the ideal
tools for such delicate work. As the rock came away, we discovered that behind
the tooth lay another tooth, and something more behind that. Perhaps, we
thought with growing excitement, there might even be an entire jaw and skull.
Leakey & van Lawick, 1963

Whether Mary Leakey’s find yet constituted evidence, evidence for a new kind of hominin,
remained to be established. And the potential constitution of the find as evidence rested
itself on other work. That anyone should bother to look here, in this area, at all for our
human forebears rested on lucky accident and intuition rather than careful pre-planning or
strong prima facie evidence. The 1931 team of palaeoanthropologists led by Louis and
Mary Leakey was drawn to the area by the earlier work of entomologists and geologists
who had by chance, as part of their insect and rock hunting, unearthed a human skeleton.
This was thought at first to be an ancient fossil, but only much later was discovered to be
a relatively modern intrusive burial in older deposits. Its constitution as evidence of the
presence of fossil Hominoidea, drawing palaeoanthropologists to the area like a magnet,
was thus mistaken. And the Leakeys’ subsequent moves around the area were
determined as much by what is often called ‘intuition’ as ‘evidence’. As Louis Leakey
himself put it, ‘For some reason both of us [he and his wife] had been drawn again and
again to this particular site’ (Leakey & van Lawick, 1963: 134).
It is the ‘For some reason’ that is primarily of interest here, for it gives a clue about what is meant by ‘intuition’, a meaning residing in personal, tacit knowledge built out of information – data, evidence – accumulated both deliberately and fortuitously about the world. Ideas often emerge out of confluences of circumstantial evidence in the minds of those who are steeped in a problem, a discourse or a technology. Often those individuals have a feeling, a hunch that this way or that way is the right way to proceed, without being able to articulate its evidential provenance. There is a playing around with bits and pieces of everyday evidence that in some way enables practitioners of one kind or another to discover – in the words often heard nowadays in relation to evidence-based practice – ‘what works’. The discovery by the Wright brothers that the aeroplane worked (see Dyson, 1997: 17) was made outside the theoretical frameworks within which evidence is supposedly used and from which it supposedly emerges. The evidence behind the design of their aeroplane came from observation, trial and error, and from noting the trials and tribulations of others; the evidence of its success came from the fact that it flew – no more and no less. No randomised controlled trials were necessary.

The discovery of penicillin, the invention of nylon, the discovery of superconductivity (see de Bruyn Ouboter, 1996) are all well documented cases of the ‘intelligent noticing’ of evidence that emerged outside the intellectual infrastructure from which evidence is expected to materialise. Clearly, all were discovered by able people working with the tools of their trade and immersed in the ideas of their intellectual communities. But their significant breakthroughs – and this applies as much to the everyday insight in practical activity as to the important breakthrough – occurred out of evidence collected almost incidentally and worked on with personal knowledge and the knowledge of the intellectual communities of which they were a part.

All scientists – whether physicists, chemists, biologists, palaeoanthropologists – use particular kinds of evidence and meld it in particular ways relevant to their fields of work and the methodological traditions that have developed there.

The interesting thing about scientists’ discourse in reflecting on these methodological traditions is that there is generally accepted to be no particular, no correct or proper way of generating or marshalling evidence. As Einstein put it, the creative scientist must be an ‘unscrupulous opportunist’. The essence of science, he said, is the seeking ‘in whatever manner is suitable, a simplified and lucid image of the world … There is no logical path, but only intuition’ (cited in Holton, 1995: 168). In a similar vein, Feyerabend (1993: 14) asserted that thought actually moves forward by ‘a maze of interactions … by accidents and conjunctures and curious juxtapositions of events.’

This use of the word ‘intuition’ by Einstein in the context of scientific endeavour is an interesting one. Clearly he is not denying the significance of evidence. Rather, he seems to be promoting a kind of spatchcock use of evidence – a playing with almost any piece of seemingly relevant information – by scientists in their everyday work. And here scientists are very much like teachers or doctors, operating on their own and as part of a community, and drawing eclectically from many and varied streams of information: a bricloage of potential evidence.

Leakey, for example, describes how what could have been mere informational noise became evidence. First of all came the pieces of supporting evidence to complete a jigsaw:

… after several days we had all the pieces out and began putting our fossil jigsaw puzzle together. At last we could see what we had: Mary had discovered a nearly complete skull of Proconsul africanus, an early Miocene creature.

Then came corroborative evidence, of the fossil’s age from potassium argon dating, and confirmation of its significance from other experts in the field. The presence of a canine
fossa, for example, is accepted in the scientific community as evidence of presence on the branch to *Hominoidea* – the canine serving, as it does, as an anchor for a muscle which controls the movement of the upper lip, and therefore probably being important in some form of proto-speech.

The example of palaeoanthropology is given here to address some common themes in the search for evidence. In palaeoanthropology and elsewhere, in order for information to constitute evidence, that information has to pass a number of tests.

**First, its relevance has to be determined**, for the notion of evidence assumes that something less than established fact – an *assertion*, a proposition, an hypothesis – has been put forward and data of some kind is wanted in support of that position. The data inscribed on a music CD is not in itself evidence for anything unless assertions are made about it (for example, that the music is good, in which case the data would constitute evidence about which judgements would have to be made). Evidence is thus information supporting (or refuting) an assertion, and must pass the test of relevance if it is to move from informational noise, to potential evidence through to *prima facie* evidence.

In helping to determine relevance will be other pieces of information, and this raises the second test – of **sufficiency**: the potential evidence has to be considered with other pieces of information to determine its place in support of the assertion. Is there, in other words, *corroborating evidence*? As Russell points out, the person who asserts that unsupported bodies in air fall, has merely generalised from insufficient evidence, ‘… and is liable to be refuted by balloons, butterflies and aeroplanes’ (1956: 91). Quality and sufficiency of evidence here begin to be related to the epistemological ambitions of the proposition.

Third, and again linked to sufficiency, decisions have to be made about the **veracity** of the evidence: were these, in the Leakeys’ example, more modern ‘intrusions’, and if so what was their status – were they perhaps *fakes*, as had been the case with Piltdown Man? Or, are commitments so intense that the strong possibility of selectivity and misinterpretation make the ‘evidence base’ untrustworthy? There are cases in the education archive of commitment being so powerful that deliberate manipulation and distortion of evidence can be postulated. Such questions arise principally because of the interests that exist in any research enterprise – interests usually surrounding personal gain of one kind or another. Here, the existence of corroborative evidence is again essential in helping to determine veracity.

These various tests for evidence are summarised in Table 1.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Enabled by</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. relevance</td>
<td>establishing that the information constitutes information for (or against) some proposition</td>
</tr>
<tr>
<td>2. sufficiency</td>
<td>corroboration with other instances of the same kind of evidence or other kinds of evidence</td>
</tr>
<tr>
<td>3. veracity</td>
<td>establishing that the process of gathering evidence has been free from distortion and as far as possible uncontaminated by vested interest</td>
</tr>
</tbody>
</table>

There is a second reason for using palaeoanthropology as an example, and that is to demonstrate how notions of evidence and of research practice can vary across different
domains of inquiry and practice. One is struck by the fact that there appears to be little role for experimentation in this respected science: there is a reliance on idiographic evidence with a trust in the testimony of expert peers and commentators, and large doses of trust in those who made the find. Replication of the work, in any meaningful sense, is not feasible.

Perhaps here, in examining the routes taken by different communities of inquiry in establishing evidence, one can borrow from the distinction that Lévi-Strauss draws between *bricoleurs* and *engineers* (see Norris, 1987: 133-34). The bricoleur begins with a broad proposition and collects evidence *ad hoc*, within the broadest parameters and using the least delimiting rules, expecting to discover useful *prima facie* information. This *prima facie* evidence is examined for its veracity while additional corroborative evidence is sought. This seems to reflect the practice in the Leakeys’ work, where clues led to finds which were then verified with corroborative evidence. The engineer, by contrast, is assumed to be bound initially within strict theoretical frameworks, beginning with a clear view of the object to be constructed, and ultimately developing a blueprint that will be closely followed.

In fact, as both Lévi-Strauss and Derrida (Derrida, 1978: 285) note, processes are not nearly as cleanly distinct as the metaphors of bricoleur and engineer imply. In real life, engineers and physicists, as much as anthropologists, cast the net far and wide in their search for evidence. Wright Mills (1970) quotes a number of Nobel-Prize-winning physicists on the process of gathering evidence and using it to establish reliable knowledge.

Bridgman, for example, says that ‘There is no scientific method as such, but the vital feature of the scientist’s procedure has been merely to do his utmost with his mind, *no holds barred*’ [original emphasis] and Beck says that ‘The mechanics of discovery are not known … I think that the creative process is so closely tied in with the emotional structure of an individual … that … it is a poor subject for generalisation’ (Wright Mills 1970: 69).

A problem for scientists in discussing evidence and its use is similar to that faced by teachers and educational researchers: of being seen in one’s public face to be rigorous, methodical and meticulous; of creating and maintaining an impression of the kind of systematic study assumed by the general public to be the hallmark of science. The problem was well expressed by the renowned biologist Peter Medawar, who – in discussing scientific method – talked of ‘the postures we choose to be seen in when the curtain goes up and the public sees us’ (Medawar, 1982: 88). The danger, of course, lies in playing with bits and pieces of random evidence: ‘Most [discoveries] entered the mind by processes of the kind vaguely called “intuitive” … this is seldom apparent from mathematical writings because mathematicians take pains to ensure that it should not be’ (ibid: 87-8).

***

It is worth pausing here to look at philosophers’ ‘take’ on evidence, since they perhaps consider evidence *per se* more than scientists or educators. For them, evidence constitutes any information that bears on the truth or falsity of a proposition. One should note again that *sufficiency* is important also for the philosopher in determining the status of one’s evidence in moving beyond information, and in particular for drawing distinctions between the status of the beliefs one holds. Many age-old philosophical debates, such as those about our knowledge of the external world, or the basis for moral judgements, are largely about whether the evidence we have in these areas is sufficient to yield knowledge – or merely rational belief. Is the evidence weak, strong … conclusive?
Rational belief is perhaps all that can be hoped for in practical circumstances, and it is unlikely that a practitioner will find conclusive evidence for a proposition, since conclusive evidence is so strong as to rule out any possibility of error. Conclusive evidence is always going to be lacking for beliefs about the external world, whether those beliefs concern the past, other minds, or the efficacy of Reading Recovery.

Aside from the significance of sufficiency, another strand has to be drawn out of this overview and this is the social and interpretative context of evidence. In any community of inquiry – scientific, legal, artistic – both the collection of evidence and its assessment is part of a social process. In jurisprudence, for example, as in science, a premium is put on the process of corroboration for establishing both the relevance and valency of evidence. For lawyers, evidence constitutes the means by which disputed facts are proved to be true or untrue and in English law the principal criterion determining the acceptability or admissibility of evidence is when it is relevant, that is to say when it has a tendency ‘in reason’ to prove or disprove disputed facts. And the employment of ‘reason’ to determine the relevance, value or effectiveness of evidence necessarily makes the process of the acceptance of potential evidence a social one, depending on how persuasive the evidence seems, for example to a jury. There is nothing inherently objective or acceptable in a particular piece of evidence: nothing to guarantee its verisimilitude. It may be an eyewitness testimony, a DNA fingerprint or a document but its value depends not so much on its placing in the canon of good sources but rather on its position in the panoply of factors – kinds of evidence – that surround its discovery. And the process by which the assessment is made of that contextual, corroborative evidence is a social one. Who says, and who is prepared to accept, that this evidence is sound?

There is therefore a similarity in the ways in which the legal and the scientific communities approach the notion of evidence. In each, the ultimate determination of the value of the evidence will rest on the judgement of peers – those peers being twelve ordinary people in the case of the jury, and an expert scientific community (who will replicate, convene, confer, ‘peer-review’ and judge) in the case of science. Both in the law and in science this social process will determine the quality and sufficiency of the corroborative evidence: has enough independent, admissible evidence confirmed, for example, that the accused has committed the crime? Or does one piece of very strong evidence outweigh several pieces of seemingly flimsier evidence? It is the ‘seemingly’ that is interesting here. The key matter is how the evidence seems to a community of assessors. The process of assessment is social.

And as soon as the process becomes social, questions arise about the interpenetration of notions that have travelled along different paths to the same question. Even for natural scientists, political is mixed with empirical for some of the evidential matters that they confront. As biologists Levins and Lewontin (1985: 4) point out, ‘The denial of the interpenetration of the scientific and the social is itself a political act, giving support to social structures that hide behind scientific objectivity’. Levins and Lewontin go on to note that ‘Of course the speed of light is the same under socialism and capitalism’ but that social questions – questions that arise about the way people live – cannot be answered without reference to an interrelated range of matters. Thus, the experimental evidence will suggest that the cause of tuberculosis is a bacillus, yet the evidence of epidemiologists tells us that the disease rarely takes hold where people’s living and working conditions are adequate. In education, matters are no less complex.

***
Evidence and ‘evidence-based practice’

This social context of evidence is crucial in determining its validity. I wish now however to return to the notion of sufficiency, for it is around the need for sufficiency – around the need for collection and collation of good quality, reliable evidence – that arguments in particular for the employment of ‘evidence-based practice’ in education usually pivot. If a theme can be distilled from the discussion thus far in relation to sufficiency, it is that in strengthening a belief there must be a movement beyond the first piece of evidence to additional corroborative evidence (see Figure 1).

Figure 1: A continuum of sufficiency

<table>
<thead>
<tr>
<th>Kinds of evidence</th>
<th>isolated observations</th>
<th>prima facie or inconclusive evidence</th>
<th>corroborative evidence</th>
<th>conclusive evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leading to</td>
<td>inspiration</td>
<td>hunch</td>
<td>rational belief</td>
<td>knowledge</td>
</tr>
</tbody>
</table>

Questions arise, however, about the meaningfulness of such a typology when talking of evidence for practical activity and the beliefs inhering therein, for movement along the continuum in Figure 1 appears to be a deliberate, calculated process. Yet Polanyi (1969) reminds us that the scientist, for example, cannot specify the way in which facts become evidence. Neither, in dealing with the world more generally is the process of gathering evidence and placing it in a theoretical or heuristic framework for practical purposes a conscious one. Polanyi (1969: 144) notes:

> We cannot learn to keep our balance on a bicycle by taking to heart that in order to compensate for a given angle of imbalance \( \alpha \), we must take a curve on the side of the imbalance, of which the radius \( r \) should be proportionate to the square of 5\( \text{th} \) the velocity \( v \) over the imbalance: \( r \sim v^2 / \alpha \). Such knowledge is ineffectual, unless known tacitly.

The same kind of thing could be said of much professional knowledge in teaching. That craft knowledge in the classroom – whether to make eye contact, how to respond to an interruption, what sort of question to ask, what kind of language to employ in developing a theme – is often ineffectual unless known tacitly.

The issue, however, for some of the proponents of evidence-based practice is not in recognising the significance of this kind of tacit knowledge but rather in understanding practitioners’ ability to reconcile it and meld it with knowledge from research: research evidence. Hargreaves (1996), for example, suggests that while medics achieve a good balance between craft knowledge and declarative research knowledge, teachers have been less successful in employing research evidence – in part because of the nature and presentation of that research evidence – alongside their craft knowledge: less successful in employing this additional corroborative evidence.

What Hargreaves seems to be suggesting is outlined in Figure 2: a more systematic incorporation of research evidence to the tacit knowledge/craft knowledge \( \rightarrow \) practice cycle. This is essentially the position of many of those who promote evidence-based practice in education.
One must ask questions, though, about the status of the declarative knowledge and the kind of evidence needed to achieve it in different domains of practice. How valid is it to make the assertion that this evidence, in the practical circumstance of the teacher – as distinct from the medic – is corroborative?

***

In particular, one needs to examine the meaning of evidence-based practice among its proponents in education, for it is here especially that questions arise about what is likely to be gained in practical circumstances from particular kinds of evidence.

If various stages in the employment of evidence are traversed in moving toward knowledge – a bricolage/hunch stage, an inspirational stage, a discovery stage and a corroborative/confirmatory stage – the notion of evidence-based practice focuses on evidence at the confirmatory stage, on the systematic collation of research studies for use by practitioners and policy makers.

Oakley (2000) speaks specifically about this systematic collation. Drawing on a definition of evidence-based education that suggests that the establishment of sound research evidence must attend to the criteria of scientific validity, high quality and practical relevance, she suggests that systematicity is at the core of any gathering of such evidence:

Systematic reviews are the primary method for managing knowledge in the evidence movement approach. This is because they synthesise the findings of many different research studies in a way which is explicit, transparent, replicable, accountable and (potentially) updateable.

Oakley (2000: 3)

Many proponents go further, to suggest that randomised controlled trials (RCTs) are at the core of the evidential armoury and that systematicity should centre around the collection of this kind of study. Here, ‘evidence’ means a particular kind – a superior kind – of research evidence. Take what Robert E. Slavin had to say in the ‘Distinguished Lecture’ of the American Educational Research Association at its annual meeting in 2002:

The most important reason for the extraordinary advances in medicine, agriculture, and other fields is the acceptance by practitioners of evidence as the basis for practice. In particular, it is the randomized clinical trial – more than any single medical breakthrough – that has transformed medicine (Slavin, 2002: 16).

Figure 3 summarises the ‘robust’ position of proponents of evidence-based practice, and the one that Slavin appears to take, drawing a central distinction between evidence from research, and evidence from personal or professional experience. Research evidence is itself delineated: nomothetic, idiographic; well designed, poorly designed. Certainly varied kinds of evidence can exist and they all have a degree of validity, it is conceded by Slavin, but trustworthy evidence, he seems to suggest, is obtained by following the heavy line that I have drawn in Figure 3.
Figure 3: Evidence as viewed in 'evidence-based practice'

- Testimony
- Evidence from documents or archives
- Artefacts
- Personal experience
- Observation
- Research
- Idiographic research
- Poorly designed nomothetic research
- Nomothetic research
- Well designed nomothetic research, especially RCTs
Researchers and practitioners in other successful areas of professional inquiry, notably medicine, have followed it, and so should researchers and practitioners in education.

Slavin goes on to note that such experiments are in fact rare in education, concluding that ‘… the experiment is the design of choice for studies that seek to make causal connections, and particularly for evaluations of educational innovations’ (ibid: 18). Goldstein (2002) agrees that the randomised controlled trial (RCT) is generally accepted as a gold standard in applied statistical work, but in fact concludes that its use does not enable the establishment of causal connections. Nor is its use necessary for causal connections to be inferred.

It should also be said in relation to Slavin’s assertion that although there have clearly been ‘extraordinary advances’ in medicine and agriculture in recent years, it is not clear (to me at least) how these can be attributed to the benefits of the RCT. The RCT plays an important but mundane, confirmatory role; it cannot provide the research work that leads to the advance. The recent remarkable discovery, for example, that the bacterium *Helicobacter pylori* causes more than 90 per cent of all peptic ulcers, came not from RCTs but from (as Medawar, above, would have predicted) a mix of inspirational thinking and serendipitous events (see Blaser, 1996). RCTs merely confirmed the efficacy of antibiotic therapy in the killing of the bacteria. Once the discovery had been made that bacteria were the culprits (by the ‘Aha!’ of an observant expert, not by a RCT) it required not a huge cognitive leap to get to the point that the bacteria needed killing in ulcer patients’ stomachs. RCTs played a dull but important role in confirming the effectiveness of antibiotics in doing the killing. The point is that experiments and RCTs do not, as Slavin suggests, lead to ‘extraordinary advances’.

Indeed, even in the RCT’s workings as confirmatory agent, it may lead to falsely high or low expectations about the expected benefits of the change, and as Matthews (1998:17) puts it ‘… can easily double the real efficacy of a useless drug’ (see also Matthews, 2001). This phenomenon, generated by the failure of frequentist methods to take into account *prior evidence* has of late led to calls for methods which contextualise *new evidence* in the context of *existing evidence* (see, for example, Spiegelhalter, 1999; Howard et al, 2000).

Caveats notwithstanding, though, such experiments and their analysis are taken to be central to the building of an evidence base for practice. Some time ago, Parlett and Hamilton (1987) described this kind of work as emerging from an *agricultural-botany* paradigm and pointed out some of the weaknesses of its adoption in education, and there have been many critiques both of this framework and the frequentist methods behind it in the social sciences (eg Cohen, 1994; Howard et al, 2000). The particular criticisms of its use in education made by Parlett and Hamilton centred on assumptions made about the stability of ‘before’ and ‘after’ in before and after research designs, and about assumptions about the robustness of, for example, test scores and attitude ratings (often used to indicate ‘effectiveness’ of one kind or another) – which are assumed to be of the same order of robustness as the measures made, for example, of plant growth. In short, they argued, study undertaken in education in this tradition falls short of its own claims to be controlled, exact and unambiguous.

***

Dialogue, of course, continues about the need for and value of such research, with questions about both its technical capacity to deliver (questions that can be met with the answer that technical improvements in the instruments will be made), and about the nature of the evidence provided: does this kind of evidence – can this kind of evidence – enhance practice? Teachers come to teaching with a set of beliefs and understandings
and these sometimes seem to be impermeable (or at least only semi-permeable) to the kinds of evidence emerging from certain kinds of educational research. One of the interesting aspects of teachers’ practice is its seeming resistance to evidence of particular kinds and its solidarity in a certain commonality of purpose, which can perhaps be located in what Oakeshott (1967: 157) called ‘an inheritance of feelings, emotions, images, visions, thoughts, beliefs, ideas, understandings’. It may be that certain kinds of research evidence – particularly if these are perceived in any way to be imposed by authority – do not mesh easily with this sort of inheritance, a problem noted by Peile (2004) to exist also amongst medics.

Practitioners accumulate evidence in practice and distil it in everyday heuristic, knack and rule of thumb. They engage in, and have confidence in, a kind of vernacular accumulation of evidence that enables what Schatzman (1991: 304) calls ‘common interpretive acts’. As Schatzman suggests, we are all – as professionals and as laypersons – using these interpretive acts all the time and unselfconsciously employing them to help us order and comprehend the world. We all find pieces of evidence, make links between them, discover patterns, make generalisations, create explanatory propositions all the time, emerging out of our experience, and this is all ‘empirical’. All professionals will collect evidence deliberately and tacitly in ways described eloquently by Schön (1991) and others, who emphasise the interconnectedness of professional knowledge. The evidence will be reviewed, talked about with colleagues, new things will be tried out as a consequence and informally evaluated. Practitioners’ trust in the knowledge that such processes work is perhaps at the root of their resistance to the imposition of other kinds of evidence.

References


