

# The uses and abuses of evidence in education

Geoff Petty April 2015 draft 2

---

## The best evidence is flawed, the rest is worse. But there are ways to navigate this uncertainty

### Summary

No evidence or advice is perfect, but some sources of evidence are much more trustworthy than others. You can often tell a better source from a worse one, as I will explain. High quality evidence will suggest many ways to help you improve your practice, and steer you away from approaches that don't work so well. This will save you much time and trouble.

However, if a teaching method<sup>1</sup> works in theory, or in other people's classrooms, that's no guarantee it will work for you. The ultimate authority is your own professional experience: does the method work for you and your students? Don't, though, abandon a method if it doesn't work first time. Because of the great complexity in teaching, it usually takes a lot of trial and error to get a method working well, even for experienced teachers.

Any teacher can become an outstanding teacher, if they learn well how to use outstanding teaching methods. This requires you to repeatedly trial evidence-based methods until you are using them well. If they don't work after a few trials you abandon them and trial another. It's vital you discuss this experimentation within a team of teachers - Timperley (2008).

But how do they find the methods most likely to work amongst literally tonnes of suggestions in research? I argue we should look at summary evidence from three different schools of inquiry, looking for what they have in common: qualitative research, quantitative research, and field studies on the best teachers from a value added perspective. This is called 'triangulation'.

These should tell us what methods are best to experiment with. When we have trialled a new method a few times in our own classrooms we should trust our own professional judgement as to whether we can make the method work for us, and our students.

When I read how to improve teaching, authors hardly ever use triangulation, and bias is common.

### Sifting the evidence: the Bias Test

To find the best evidence and advice we must overcome bias – in ourselves and in others. Bias includes:

*Confirmation bias*: We tend to look for, choose and remember ideas that confirm our present views and avoid or forget those that contradict them. Lefties read the Guardian, Righties the Telegraph. It's just easier to stick with your present views and practices.

*Groupthink*: If all my colleagues think ability grouping is a good idea, I will feel uncomfortable disagreeing with them. We are tribal animals, and we like to fit in.

*Doc shopping*: Given the above there is a strong tendency to look for and quote authorities who agree with us. It is not hard to find **doctorates** (Phds) and **documents** on the web that reflect your prejudices with persuasive

---

<sup>1</sup> By 'method' I mean a teaching strategy, approach, technique, resource etc - small or large.

eloquence. But most docs might *disagree* with you! It's not just you who has this problem, everyone you read has it too, to some degree.

Things are looking complicated! How do we steer a true course through this ocean of self-deluding opinion? Who can we trust? Lets look critically at a few sources of evidence and see which overcome the biases mentioned above, and so pass 'the bias test'.

## Sources of Evidence

*'Well it works for me'*: Okay, but could something else work even better? Teaching is impossible to do perfectly and there are lots of alternative approaches.

*'Everybody else does it'*: This is groupthink. Sometimes groups are right, sometimes not. The most commonly used questioning methods in uk classrooms are among the worst available.<sup>2</sup>

*'What Ofsted recommends'*: Ofsted does not dictate *how* you should teach, thankfully. The inspection handbook, published 31st July 2014, states that:

*"Ofsted does not favour any particular teaching style and inspectors must not give the impression that it does. School leaders and teachers should decide for themselves how best to teach, and be given the opportunity, through questioning by inspectors, to explain why they have made the decisions they have and provide evidence of the effectiveness of their choices."*

Ofsted wants outcomes, not particular methods, so you are on your own. Be grateful! However, many people claim to know what methods Ofsted inspectors are 'really' looking for – ask them how they know. I've never heard a credible answer.

Ofsted publishes advice on 'best practice' but it's only based on their own experience and so doesn't pass the bias test. (Christodoulou (2014) criticises their advice on best practice, but see my blog on her book which has other weaknesses.)

*Read published research studies* : An advantage of anything published in a recognised journal is that it will normally be 'peer reviewed' meaning some anonymous experts will have vetted the study. If the study was not published it probably hasn't been vetted.

Individual studies can often provide the very information you are after though you need good searching skills, ask a librarian to help you.

A problem with reading individual studies is that for every study that says one thing, there may be others that say the opposite. And you don't have time to read *all* the studies on any given issue or method. Luckily someone else may have done that work for you and more – to create a 'research review' considered below.

*Read published quantitative research studies* : The only way of finding out whether something works is to try it out with a real teacher and real students in a rigorous trial. The best trials randomly split the students into two very similar groups, the 'experimental group' experiences the method, the very similar 'control group' gets conventional teaching instead. The researchers then test the students to see if they learn better with the method (experimental group) than without it (control group). Researchers can then calculate the improvement in learning brought about by the method using the unit of 'effect size'. An effect size of 0.4 is average, 0.6 is high, 0.2 is a very small improvement in learning, (Hattie 2009).

Effect size is not an entirely reliable measure<sup>3</sup>. However it is the only means we have of comparing one teaching strategy with another to see which has been most effective in other people's classrooms. Without such a comparison

---

<sup>2</sup> <http://geoffpetty.com/for-teachers/questioning/> Petty, G. (2009) 'Evidence Based Teaching' chap 15

<sup>3</sup> [Why education will never be a research-based profession](http://www.dylanwilliam.org/Dylan_Wiliams_website/Presentations.html)  
[www.dylanwilliam.org/Dylan\\_Wiliams\\_website/Presentations.html](http://www.dylanwilliam.org/Dylan_Wiliams_website/Presentations.html)

we have *no* idea whether we should experiment with, say, self-assessment or team teaching. Some methods are very powerful, we need to know which ones. Critics of effect size must explain why some methods repeatedly get very high effect sizes in hundreds of rigorous studies, and when these methods are well trialled in classrooms, teachers eventually get corresponding boosts in student attainment<sup>4</sup>.

One problem is that the research may have been carried out in a different context to your own teaching, for example in secondary school rather than in a college. However this context often does not make much difference, (research reviews test for this in any case).

*Read research reviews* : These include ‘meta-studies’, ‘meta-analyses’ ‘Best Evidence Syntheses’ and systematic reviews. These are done by experts who systematically look at **all** research on a given topic, say formative assessment. They reject poor quality studies that don’t meet objective criteria such as ‘the study must involve more than 30 students’. It’s common for the vast majority of studies to be rejected. The reviewers are not allowed to reject studies because they don’t agree with them!

Having found the best quality studies, the reviewers consider them all together and summarise what they show, pointing out agreements and disagreements in the research. This approach minimises the biases mentioned earlier. It also does a lot of reading for you!

An example of a research review is Nesbit (2006) on learning with concept and knowledge maps (mind-mapping). The studies in his review showed that mind-mapping *could* create an effect size of about 1.0. Students were then learning at twice the rate of the very similar students in the control group who did something else instead of mapping. Nesbit discovers which uses of these maps gives the highest effect-sizes, which is really useful. For example he finds it helps students understand and remember central ideas more than it helps them with detail. Good reviews suggest how to use methods effectively in your classroom, at least in outline. Some are vague about this practical detail though.

There are more than a thousand research reviews on factors that affect student achievement, this sounds a lot, but they might not answer your specific question, whereas an individual study might.

*Read cognitive psychology research studies* : These give the theory of how to teach based on experiments which are mainly done in psychology labs. An example is that if students’ present understandings of a sub-topic are checked and corrected, they will better understand new concepts and ideas in that same sub-topic. This is very helpful, but detailed advice on **how** to do this will not usually have been tried out in a classroom by cognitive psychologists. (However, in this case quantitative research confirms a strategy based on this idea: ‘*Relevant Recall Questions*’ and it has a high effect size (Marzano 2001)

Research reviews are more useful than reading individual studies usually, as they consider all the evidence dispassionately. One major review of research in cognitive psychology is Bransford (2000).

*Read books by experts* : Experts are subject to bias unless they make fair use of the best evidence available. Books are edited, but they are rarely peer reviewed before publication. If a book is recommended you might think it will be a more reliable or useful guide. But does the recommender pass the bias test, or do they like the book because it confirms their preferred practice and prejudices?

*Advice backed up by references* : People often give references to back up their opinions. But if all their references are just individual studies, then they may well be omitting the evidence **against** their point of view. If they reference research **reviews** this is much more reliable. Look for terms like this in their references: ‘meta-study’, ‘meta-analysis’ ‘Best Evidence Syntheses (BES)’ or ‘systematic review’.

*Read blogs, websites, twitter, newspaper articles etc* : Are the ideas based on evidence in research reviews, or just individual studies, or on opinion only? A favourite trick is to find a poor piece of research, or an extreme view, and correctly criticise this, leaving gullible readers believing the author has disproved that position, and so proved their own. Neither was the case. Misrepresenting an opponent’s view and then arguing against it is called a ‘straw man argument’.

---

search also for ‘Randomized control trials in education research Dylan Wiliam’

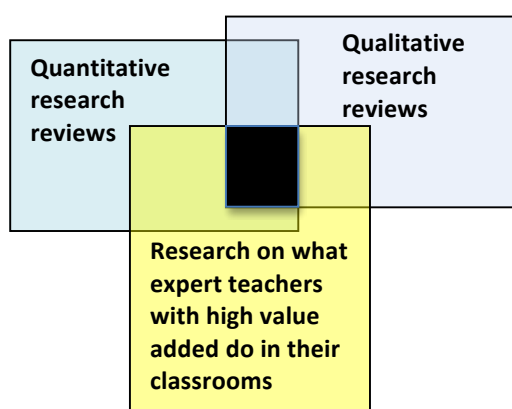
<sup>4</sup> Hattie (2009). Dylan Wiliam (2009)

Many social media and newspaper articles are thought provoking, informative, up-to-the-minute, and useful but systematic reviews are more trustworthy.

*Read research on what the best value-added teachers do* : There is very little of this type of research and even less of it reviewed, but the findings are going to be more relevant than individual quantitative studies, as the teachers’ performance was evaluated over a much longer time than in most effect-size studies. We need much more of this type of research.

I’m deliberately setting a very high standard here. Much government advice does not pass these tests. But teaching is important: it can change lives. We need to be careful. When trainee teachers have followed the above advice for a bit and have some experience with it, they will be able to trust their own opinion more. A table at the end of the paper summarises the strengths of the above sources of evidence.

## Using more than one source of evidence: ‘triangulation’



The overlap between these research fields is larger than my diagram suggests. Also, many quantitative reviews include qualitative findings, so the sources of evidence are not as separate as the diagram suggests.

	Some advantages	Some disadvantages
<b>Quantitative research reviews</b>	Allow us to compare effect sizes and so prioritise what to experiment with. Real classrooms and students.	Effect size measures are not entirely reliable. Only achievement considered usually.
<b>Qualitative research reviews</b>	Give us theoretical understanding of the learning process. Helps explain why methods work.	Research done in a laboratory context, not in classrooms. Can’t compare methods.
<b>Research on what expert teachers with high value-added do in their classrooms. “Field research”</b>	The teachers have exceptional achievement over many years rather than just during a study. Real classrooms and students.	There is very little of this research

The table above greatly simplifies a very complex situation on the best forms of evidence. Notice that no source of evidence gets unqualified approval, they all make a useful contribution, and they all have weaknesses. Educationalists sometimes critique one of these schools of evidence, ignoring its strengths, while ignoring the weaknesses of the alternatives they advocate. This selective perfectionism is unfair, common on the web, and not uncommon in books and journals.

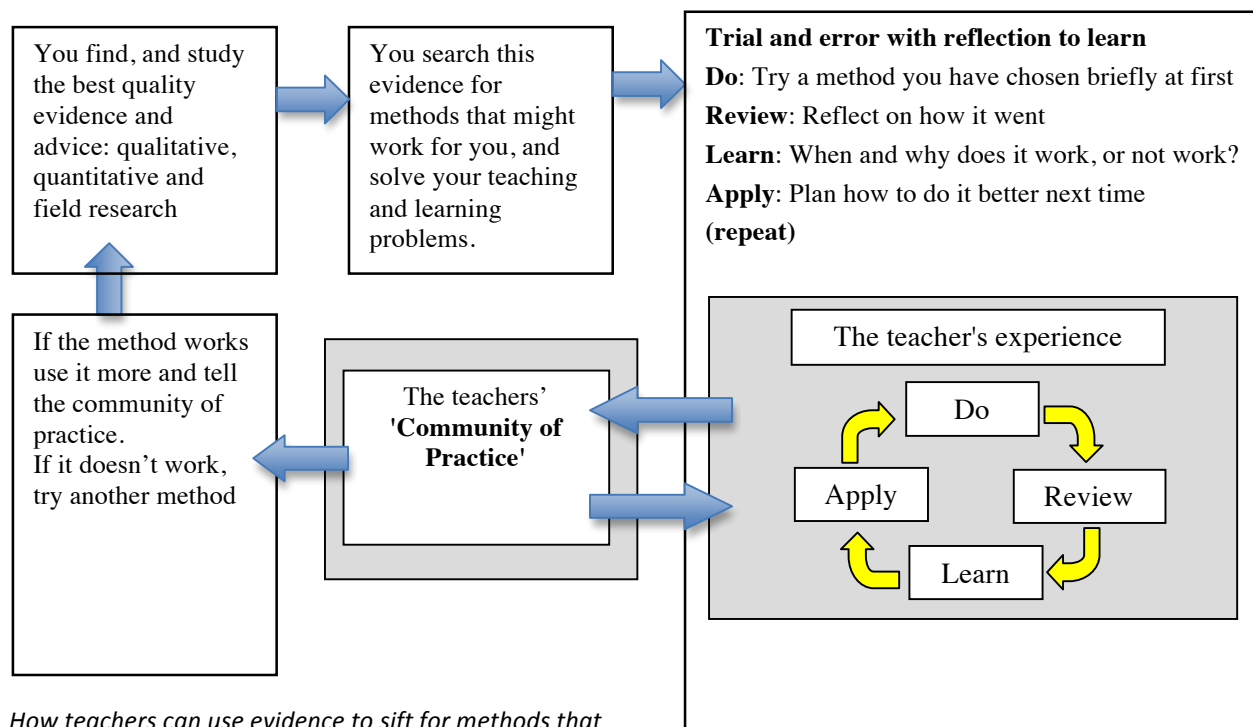
I suggest we do what journalists are taught to do: use multiple sources of evidence. If a method, strategy or other variable is recommended by qualitative research, has a high effect size, and is used by teachers who get exceptional value-added, then it is worth a try in our own classrooms. Especially if it might fix a problem we, or our learners are having. If a method gets the thumbs up from two out of three of these schools it is also worth consideration.

I used this ‘triangulation’ approach in ‘Evidence Based Teaching’, Petty (2009). Triangulation is a widely accepted approach used in many areas, for example the Intergovernmental Panel on Climate Change uses something like it, though much more nuanced. In education I believe something like it is used by the Education Endowment Fund, and

by the Eppi-Centre, but again much more nuanced. Dylan Wiliam one of education's most respected researcher and research reviewer, writes that [effect-size studies] "benefit from research designs that include complementary approaches to inquiry". He adds that criticisms of effect size studies do not mean that "they are a bad idea" though if we rely exclusively upon them "we end up not being able to say very much." Wiliam (2014). I take him to mean we should use multiple sources of evidence.

However, some organisations which claim to use evidence or research are much more casual and cavalier in their use of research and evidence. Triangulation is a minimum expectation if we are serious about the use of research.

## Making use of the evidence to improve your teaching



*How teachers can use evidence to sift for methods that might improve their teaching*

The diagram shows how high quality evidence can be used to suggest methods that will work for you and solve your teaching and learning problems. Note that the evidence is not your dictator, only your advisor. It need not limit your teaching or creativity in any way. You choose a useful looking method and experiment with it repeatedly, learning how to use it well (this is sometimes called 'action research'). You discuss your experiments with a group of teachers who are also experimenting with their methods. This is called a 'Community of Practice': research reviews on how to improve teaching show Communities of Practice are vitally necessary for your own improvement<sup>5</sup>.

If the method you are experimenting with works, great, you use it more. If it doesn't work after at least five trials or so, you abandon it and try another method suggested by the evidence.

Using the evidence as a source of ideas is very helpful, but shouldn't stop you thinking of your own approaches and experimenting with these. However, trialling evidence-based methods will help you understand better what works and what doesn't, so you can devise your own methods more successfully. Indeed, that's the whole purpose of using evidence; it improves your understanding of the teaching and learning process; Timperley (2007) &(2011)

<sup>5</sup> Helen Timperley et al (2007) 'Teacher professional learning and development' BES (free to download)  
B. Joyce & B. Showers (2003) Student Achievement Through Staff Development, Alexandria: ASCD. This book reviews research but does not itself seem to be systematic, or peer reviewed.

## But can you trust this paper?

It is for you to decide whether this paper is trustworthy. At least I've based the teacher improvement ideas on research reviews. But I hope it will help you devise your own ideas on where to go to get high quality evidence and advice on how to improve your teaching. This is a vital part of developing your professional practice.

If you would like a more detailed policy paper on this topic you can download one [here](#).

The following table needs to be taken with a pinch of salt. It only works for average examples of the sources of evidence. Real examples will be better or worse than this average. It would be better to use the criteria at the top of the table to evaluate the source of evidence you are considering, that would be much fairer.

## Some references on evidence based teaching

The more stars the better, a blank means no stars at all!

Comparing different sources of evidence: Warning: it works for typical evidence, but not for all cases					
	Is it checked with others, preferably peer reviewed?	Does it try systematically to include all good evidence?	Are the findings based on classrooms (preferably) like my own?	Is there a control group?	Is there an effect size to compare a method with others?
<i>What works for me'</i>			★ ★ ★ ★		
<i>What Everybody else does</i>	★ ★		★ ★ ★		
<i>What Ofsted recommends</i>	★ ★ ★		★ ★ ★		
<i>Published quantitative research study</i>	★ ★ ★		★ ★ ★	Usually not always, up to: ★ ★ ★	★ ★ ★
<i>Quantitative research reviews</i> : These include 'meta-studies', meta-analyses' 'Best Evidence Syntheses' and systematic reviews	★ ★ ★ ★	★ ★ ★	★ ★ ★	★ ★ ★	★ ★ ★ ★
<i>Qualitative research reviews</i> : e.g. systematic reviews	★ ★ ★ ★	★ ★ ★	★ ★	★	
<i>Cognitive psychology research studies</i>	★ ★		★	Not always	
<i>Books by experts</i>	★ ★	Usually not	Sometimes		Usually not
<i>Advice backed up by references</i>	★		Sometimes	Usually not	Usually not
<i>Blogs, websites, twitter etc</i>	There are often comments by others on blogs		Many blogs are		Usually not
<i>Read research on what the best value added teachers do</i>	★ ★	In effect It usually includes all teachers, but certainly not all the evidence.	★ ★ ★	★ ★ ★ The control is those teachers with lower value-added.	No, but these teachers were compared to others.

## References

- Bransford, J. D., et al. (2000) *How People Learn: brain, mind, experience and school*, Washington: National Research Council.
- Gough D, Oliver S, Thomas J (2012) *An Introduction to Systematic Reviews*. London: Sage
- Nesbit, J. C. Adesope, O. (2006) Learning with Concept and knowledge maps: a Meta-analysis. *Review of Educational Research* 76; p413
- Marzano R. Pickering, D. Pollock, J. (2001) "Classroom Instruction that works" Alexandria: ASCD
- Petty, G. (2009) *Evidence Based Teaching 2<sup>nd</sup> Ed.* Oxford University Press
- Timperley, H. et al (2008) 'Best Evidence Synthesis on Professional Learning and Development. Report to the Ministry of Education, Wellington, New Zealand
- Timperley, H. (2011) 'Realising the power of professional Learning' OUP: Maidenhead
- Wiliam, D (2009) *Assessment for learning: Why, what and how?* London: IoE
- Wiliam, D. ((2014) Randomised control trials in education research in 'Research in Education' Vol 6, No.1 University of Brighton (available on line search for title and author)

### Some interesting blogs:

#### **Andy Tharby on bias**

<https://reflectingenglish.wordpress.com/2015/04/01/the-art-of-teaching-clearly-on-why-we-should-treat-intuition-with-care/>

#### **Gary Jones on a definition of evidence informed practice for teachers and schools**

<http://evidencebasededucationalleadership.blogspot.co.uk/2015/04/the-school-research-lead-definition-of.html>

#### **Nice paper on correlations not being causes**

<http://www.huntingenglish.com/2015/03/21/causation-and-correlation-in-education/>