

Everything old is new again

“Five years ago, there was almost nothing known about how educators can use research well to improve practice.”

(Deliberately unattributed, 2024)

**Kevin
Wheldall**

**Robyn
Wheldall**



We'd like to begin by stating rather bluntly that, despite what some may profess, the science of learning (and especially the concept of evidence-based instruction) is not something new. Some of us have been arguing in this vein for many years. For example, Richard Riding and the first author (KW) launched the journal *Educational Psychology: an experimental journal of educational psychology* in 1981, arguing for 'Effective Educational Research' in their first editorial. In the early issues we included articles on Direct Instruction (DI), classroom seating arrangements, classroom behavioural interventions, effects of contextual cues on reading, Precision Teaching, Theory of Instruction, morphographic spelling, etc. It might seem like something bright and shiny that has emerged over the last five years or so, but this is not the case.

Another example is the case of Reading Recovery, recently and at long last officially discontinued in New Zealand, its country of origin. But we showed experimentally that it was of limited efficacy 30 years ago (and took a lot of heat for saying so!). One could argue similarly that the experimental evidence for the efficacy of phonics instruction and Direct Instruction has been known for years. Possibly the biggest educational experiment in history, Project Follow Through was completed in the late seventies. It was largely, and arguably deliberately, ignored. While we celebrate the new-found commitment to these approaches, we must not forget that they come from a long research tradition. This commends them even further. It is not the case that because research is not new that it is to be viewed as outdated or 'back to basics' or 'old school' – which are terms often used pejoratively.

Similarly, voices protesting that 'phonics only' is not enough are nothing new. Whole language enthusiasts have protested this for years in spite of the fact that no one could point to a source claiming the contrary. 'Phonics only' has never been recommended by anyone! The National Reading Panel report of 2000, nearly 25 years ago, emphasised that phonics was only one of the 'Five Big Ideas' of effective reading instruction.

As the battles in the reading wars draw to a close, at least for now, and it has become increasingly obvious and accepted that the Science of Reading Instruction is the victor, it is disappointing to see minor skirmishes breaking out even among the erstwhile allies; fighting over the spoils of war, perhaps! Our point here is not to expose the apparent hubris of individuals but rather to identify unedifying trends in current thought and to reassert our continuing commitment to securing our thinking on what we can learn from empirical research evidence, based on the scientific method. As the late lamented Christopher Hitchens stated ('Hitchens Razor'), and with which we concur, "what can be asserted without evidence can also be dismissed without evidence". Let's look at some examples.

One argument that has been aired recently is that not all specific reading interventions and programs necessarily need specific empirical evidence for their efficacy. If the program/intervention makes conceptual sense, it is argued, and is based soundly on the empirical research supporting its operational principles, then it can be recommended as sound instructional practice. We would demur from this assertion. It is quite possible for a program/intervention to be sound in theory but weak in practice. We cannot be certain of efficacy unless it is empirically tested using the scientific method. This may be inconvenient, but it is necessarily the case. This is the distinction between evidence-informed as against evidence, based practice because "extraordinary claims require extraordinary evidence" (the Sagan standard). As we shall argue later, there are levels of what constitutes acceptable evidence.

Similarly, there are those who argue, contrariwise, that what works in theory, the Science of Reading, will not necessarily work in practice because education is a much more complicated, 'nuanced' process than that, and we cannot control all the relevant variables. There may be some truth in this. But rather than discarding it as unworkable, this simply emphasises the need for further scientific research to identify and isolate these potentially confounding variables.

Yet another source of controversy within the Science of Reading community is the argument regarding the superiority of teaching sounds before letters as against letters before sounds. There are advocates favouring each of these approaches, but many of us, in the absence of empirical evidence to the contrary, would argue that either/or is a false dichotomy and that there is no reason why print-to-speech and speech-to-print should not both be taught together simultaneously.

So where does MultiLit stand in all of this? We reaffirm and hold fast to the need for the scientific method as the basis for understanding what works.

We often hear about the research to practice gap – that challenge of taking what the research tells us and translating it into effective classroom practice. Here are some ways that the gap from research to practice can be closed:

- growing teacher knowledge
- implementing evidence-based policy
- using tested approaches.

But how we can know what approaches we can be confident in using to help close the gap between research and practice?

Some 15 years ago we argued for a simple model of evidence for efficacy comprising five levels. But before rehearsing this we will reprise the research that we follow (and endeavor to create) in the MultiLit Research Unit, and in the MultiLit company.

- We rely (in the main) on research studies that have an empirical focus and that appear in peer-reviewed journals.
- We place our confidence in the research in proportion to the rigour of that research.
- We preference experimental research over correlational research.
- We look for replication of the research findings in subsequent research to evaluate whether consistent findings can be found.

In program design, we also look to the instructional literature for the best way to put programs together for classroom use. And where there are unanswered questions, we need to apply what we do know and align our next steps as closely as possible to approaches of proven effectiveness (an informed ‘best guess’ if you like).

Let’s just take a moment to remind ourselves about the empirical method:

- 1 Define the purpose of the research.
- 2 Explore theories and relevant literature supporting or challenging the research proposition.
- 3 Create a hypothesis (research question/s framed as a hypothesis) and determine measurement.
- 4 Specify methodology, research design and empirical data collection.
- 5 Conduct data analysis and compile results.
- 6 Draw conclusions.
- 7 And as we said previously, replication is very important.

There are a couple of important questions. First, is all evidence created equal? To that we would say a firm no. Second, how can we assess the strength of the evidence on which we seek to rely? To help us in this, back in 2007, the first author (KW) proposed a five-level scale. Using this scale helps us to weigh the evidence, and in some cases even reject it.

At Level 1, the evidence is research-based and makes conceptual sense in terms of current research and theory plus there are independent, replicated, randomised controlled trials (RCTs) providing strong evidence for specific program efficacy. This is the ‘gold standard’ to which all programs and interventions aspire, and such programs and interventions may be recommended with confidence. Unfortunately, they are very few in number.

At Level 2, the evidence is research-based and makes conceptual sense in terms of current research and theory, but the empirical evidence for specific program efficacy is more limited and may not include fully randomised controlled trials. This would count as ‘very promising’, and such programs could be recommended with reasonable confidence. It constitutes a ‘silver standard’ pending the collection of stronger evidence.

At Level 3, the evidence is research-based and makes conceptual sense in terms of current research and theory, but there is little or no empirical evidence for the specific efficacy of the program. Clearly, there is a need for supportive empirical evidence of specific program efficacy before such a program can be wholeheartedly recommended for wide application, but it may be ‘worth a try’

because it at least makes conceptual sense. In today’s parlance, this is an evidence-informed approach or program. This is the minimum basis for program recommendation and constitutes the ‘bronze standard’.

At Level 4, the quality of evidence is not research-based and makes no conceptual sense in the light of current research but may claim empirical evidence for specific program efficacy.

Such programs should not be adopted without further substantial empirical evidence for their efficacy and do not meet even the lowest standard of acceptability. Proponents of such programs should be invited to provide specific evidence, or at the very least cite supporting generic scientific research evidence or desist from making their claims. This is the ‘brass standard’. When highly polished it might, at first blush, superficially resemble gold but is soon shown not to be so, on closer examination.

At Level 5, there is no reliable research-based evidence, and it is predicated on assumptions counter to substantial scientific evidence to the contrary such that any empirical evidence offered should be viewed with considerable scepticism. Such programs should not only not be adopted, but the public should be warned that the programs are unlikely to be effective and, rather than meeting any standard, should be regarded as requiring the educational equivalent of a ‘health warning’. At best this is the ‘tin standard’.

So, we must bear in mind the evidence credentials of the approaches and programs that we use in our classrooms. Instructional time is precious, and everything must earn its keep. We need to throw out the tin cans (not recycle them!), leave the brass ornaments in the attic as they lose their lustre, provisionally settle for bronze medals in the lack of competing or better alternatives, admire and keep burnishing our silver accomplishments, while continuing to strive for gold.

*Emeritus Professor Kevin Wheldall AM
and Dr Robyn Wheldall
Joint Editors*

Reference

- Wheldall, K. (2007). Efficacy of educational programs and interventions. *Learning Difficulties Australia Bulletin*, 39(1), 3–4.