DRAFT FOR COMMENT

Styles of Research: Insight and Impact*

Hugh Burkhardt
Shell Centre and Schools of Education
Nottingham and Michigan State Universities

In this paper I look at the roles of different approaches to research in improving the performance of education systems. I compare the approaches characteristic of different traditions – the humanities, the sciences, engineering and the arts – all of which are recognisable in Education. I suggest that, if impact on the quality of education provided to most children, rather than just insight into it, is to become a primary research goal, the engineering approach needs greater emphasis in the balance of research effort, and research credit. Such a shift is likely to have other positive effects. The different characteristics of research, and the roles of strong and weak 'theory' are discussed.

Making a difference in practice

I aim to address both meanings of "in practice", that is to look at the *practical* problems of achieving improvements in teachers' classroom *practice* in helping students around the world learn more easily and effectively.

Research insights are a key input to this process but, obviously, they are not enough – unless you believe that most teachers read the research literature and, then, can and will turn its insights into substantial improvements in their practice. Neither is plausible, let alone true. What else is likely to be needed? There are standard answers.

A curriculum design will be needed, embodying policy decisions in the light of research and available resources. This design may be set down in 'standards' or other policy documents.

Curriculum materials, embodying the results of research, need to be designed and developed, so that work they well with typical teachers and their students in realistic circumstances of personnel and support. If done well, this demands more than research – notably a great deal of creative design input, and systematic refinement through feedback from trialing in classrooms.

* This paper is based on the closing talk at the ICMI Algebra Conference, University of Melbourne, December 9-14 2001, with thanks to Professor Kaye Stacey and her colleagues. As to background, the Shell Centre for Mathematical Education was established at the University of Nottingham in 1968; its mission, and its criteria for judging its success, have long been impact driven – "to work to help improve the teaching and learning of mathematics in classrooms, nationally and internationally".

Professional development support will be needed, because the range of classroom strategies and skills required for this kind of teaching is much broader than for traditional 'teacher-centred' mathematics teaching. This too may need design and systematic development of materials to support the leaders of these activities.

Pressure will surely be needed to encourage all those who need to change to do so. They are all busy people with well-grooved professional practice; few will make the effort that such changes demand without compelling reasons. Pressure is less expensive to provide than the various kinds of support just outlined, and thus attractive to politicians and the public, particularly when labelled "accountability". It follows that the assessment used must be well-aligned with the curriculum goals; otherwise, many teachers will only teach the aspects that are assessed.

Who is responsible in each education system for developing these elements? What might be done to help them? How far has the research community made its full potential contribution? What else may be needed?

Different styles of research

In education, the question "What is research?" has long raised controversy, even "paradigm wars". I want to take a broader view, looking across fields at the different meanings and traditions and asking how each can contribute to making a practical difference in Education. I shall focus particularly on the learning of mathematics and science, but some lessons are more general. Any such review will see that strength in research requires a variety of approaches, tailored to the problems in hand.

I will start from the definition of research used in the UK Research Assessment Exercise (RAE), in which all UK university departments were rated during 2001 on their work over the previous five years.

> " **'Research' for the purposes of the RAE** is to be understood as original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce and industry, as well as to the public and voluntary sectors; scholarship; the invention and generation of ideas and, images, performances and artifacts including design, where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction."

The breadth of this definition may surprise people. It arises from taking seriously four different traditions, characteristic respectively of the:

Humanities, Sciences, Engineering, and the Arts

If you look for a fundamental measure of quality in research across all these fields, it seems difficult to go beyond:

Impressing key people in your field

The balance of qualities which achieve this varies. What balance would be most beneficial for Education, and how far does it differ from current work and criteria of quality? I believe that all of these traditions have important contributions to make in Education, but that currently the balance of effort and of 'academic credit' is far from optimal. Let us look at each in turn, the nature of the activity and the forms of output, and their potential impact on students' learning in typical classrooms.

The 'humanities' approach

This is the oldest research tradition. From the RAE definition it is *"original investigation undertaken in order to gain knowledge and understanding; scholarship; the invention and generation of ideas..... where these lead to new or substantially improved insights"* Note that there is no tradition of empirical testing of the assertions made; critical appraisal, covering plausibility, internal consistency and fit to prevailing wisdoms, provides the test. The key product is **critical commentary**.

There is a lot of this in education (including this paper!). It is still the most popular and influential tradition, partly because of the general belief that anyone can play, "expert" or not. The ideas and analysis, based on authors' reflections on their experience, are often valuable. Without the requirement of further empirical testing, a great deal of ground can be covered. However, since so many plausible ideas in education have not in practice led to improved large-scale outcomes, the lack of empirical support is a weakness. How should you distinguish reliable comment from plausible speculation? This has led to a search for "evidence-based education" and the emerging dominance in the research community of the 'science approach'.

The 'science' approach

This approach to research is focussed on the development of better *insight*, of improved understanding of "how the world works", through the analysis of phenomena, the building of models which explain them, and the empirical testing of those models. In the RAE definition, it is *"original investigation undertaken in order to gain knowledge and understanding; scholarship; the invention and generation of ideas..... where these lead to new or substantially improved insights"*. Note that this is the same wording as for the humanities approach, but with a further essential requirement for empirical testing of the assertions made, which are now called hypotheses or models. The key products are:

- **assertions**
- **evidence-based arguments in support** (including evidence)
- **evidence-based responses to key questions** on the methods used

where the evidence is expected to be empirical. The common products are:

- research journal papers
- books
- conference talks

This approach is now predominant in the research in science and mathematics education. Such research provides insights, identifies problems, and suggests possibilities. However, it does not itself generate practical solutions, even on a small scale; for that, it needs to be linked to the 'engineering' approach.

The 'engineering' approach
This is directly concerned with practical *impact*, not just understanding how the world works but helping it "to work better", by systematically developing high-quality solutions to recognised practical problems. It builds on science insights, insofar as they are available, but goes beyond them. In the RAE definition is it is *"the invention and generation of ideas.... and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction"*. Again there is an essential requirement for empirical testing of the products and processes, both formatively in the development process and in evaluation. The importance of science-based insights varies from field to field, depending how far the 'theory' is an adequate basis for design. I'll come back to this. The key products are:

tools and/or processes that work well for their intended uses and users
evidence-based evaluation and justification
responses to evaluation questions

When it includes these elements, development *is* research. However, in the academic community in Education it is often undervalued – in some places only 'insight' research in the science tradition is regarded as true research currency.

Of course, definitions are man-made – social constructs that are partly arbitrary. However, the effects of lowering the status of 'educational engineering' include:

- lower standards of materials and processes, since the imaginative design and rigorous development that good engineering demands are not widely demanded;
- lower practical impact of important results of insight-focussed research, since designers feel less need to know or use the background research;
- pressure on good practitioners in universities to produce insight research papers rather than use engineering research methods to systematically improve their practice, both in its effectiveness and transferability to others.

All this leaves a hiatus between insight research and improved classroom practice which is, to say the least, unfortunate. Society's priorities for education are mainly practical – that young people should learn as effectively as possible. The perceived failures of educational research to deliver the goods in practical terms is reflected in the low levels of support for it. If politicians have a problem to solve, is their first move to call a good researcher? Not often.

This status pattern, where the pure is valued far more than the applied, is common across fields, but it is not general at any level of research. For example, the two people who have won *two* Nobel Prizes in the same field are:

- John Bardeen , the physicist, for the *transistor*, and for the *theory of superconductivity*
- Fred Sanger, the biologist, for the *3D structure of haemoglobin* (a first in this application of X-ray crystallography) and for *the procedure for sequencing DNA.*

At least two of these are engineering in approach, being "the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction". With examples like these, education need not fear for its respectability in giving equal status to engineering research. There is also a key difference in education as a field: there is currently no industry with high standards of research-based development that takes forward prototypes developed in universities, turning them into finished products. Thus there is a greater need, and responsibility, for the academic community to do this kind of work.

When built on good 'science', good 'engineering' enables us to draw much more definite conclusions about the outcomes. This has been clear in the widely recognised contributions to field from centres, for example the Freudenthal Institute, where the work has been taken beyond the small-scale study to implementation and evaluation in general school and classroom practice.

The 'arts' approach

The approach in the fine and commercial arts may be seen as related to the 'humanities' approach rather as 'engineering' is to 'science'. In the RAE definition is it is *"the invention and generation of ideas and, images, performances and artifacts including design, where these lead to new or substantially improved insights"*. I will say little about this because, though it enriches education and could do more, it is not central to my strategic concern here, which is primarily concerned with translating insights into practical impact in classrooms and school systems.

Before going into a few aspects of the argument in a bit more detail, let me stress that this paper is not a plea for the abandonment of insight-focussed 'science' research in education – it is essential, but not enough. Rather, it is an argument about balance – that there should be much more impact-focussed 'engineering' research and that it should receive comparable recognition and reward. *The different styles can and should be complementary and mutually supportive.*

Research quality: Schoenfeld's framework

In a recent paper Alan Schoenfeld has suggested three dimensions for classifying research outputs:

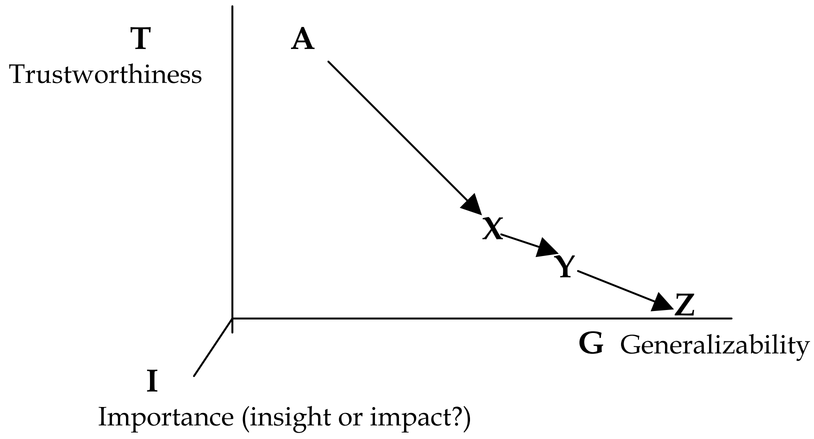
Generalizability; To how wide a set of circumstances is the statement claimed to apply?

Trustworthiness: How well substantiated are the claims?

Importance: How much should we care?

Typically, any given paper contains assertions in different parts of this 3D space of virtues. The above discussion suggests that for *Importance*, a key variable, we could usefully distinguish '*Insight*' from practical '*Impact*'.

Now I want to focus on the other two variables, G and T, say. A typical research study looks carefully at a particular situation, often a specific treatment and student responses to it. The results are high on T, low on G – the zone A on the graph below.



The conclusions section of the paper then goes on to suggest 'implications' of the study, which are often more and more wide ranging with less and less evidence to support the generalisations. These, increasingly speculative (or perhaps, experientially-basedin the humanities tradition), are illustrated as X, Y and Z.

Only large scale studies, or metanalysis, can avoid this problem. An example from the work of Alan Bell, Malcolm Swan and the Shell Centre team on 'Diagnostic Teaching' illustrates this well. (The approach is based on: leading students whose conceptual understanding is not yet robust into making errors; then getting them to understand them; and to debug their concepts and procedures through discussion) The early work showed comparable learning gains through the teaching period (pre- to post-test) but without the fall-away over the following 6-months that the comparison, a standard 'positive only' teaching approach, showed. The first study was for *one mathematics topic,* with the detailed treatment designed by *one designer*, taught by *one teacher* to *one class.* Only five years later, when the effect was shown to be stable across many topics, designers, teachers and classes could one begin to make reasonably trustworthy general statements about 'diagnostic teaching' as an approach. Even then, there remain further T2-type questions (see table below) about its accessibility to typical teachers in realistic circumstances of support, essentially: "Will ordinary teachers handle it well enough?".

The general point is that much research is really about treatments, not about the principles the authors claim to study; to probe the latter one must investigate the stability of the claims across a range of variables (student, teacher, designer and topic in this case). Typically, **this needs time and teams** beyond the scale of an individual PhD or research grant. Other subjects arrange this; if it were more common in education, the research could have high G *and* T and, if the importance were enough, be worth taking more seriously.

My final point on research styles returns briefly to the overall challenge – establishing a sound research-based path from insights to large scale implementation through good engineering. This involves work on different scales, summarized in the table below. Note the different research foci, R, and development needs, D, in the third column.

Four Levels of R&D

L	**Learning level**	*variables:* student task	*eg:* R: concepts, skills, strategies, metacognition, beliefs D: learning situations, probes, data capture
T_1	**Teaching level**	instruction student task	R: teaching tactics +strategies D: classroom materials for some teachers
T_2	**Teacher level**	teacher instruction student task	R: performance of representative teachers with realistic support D: classroom materials for most teachers
C	**System level**	system school teacher instruction student task *etc*	R: system change D: 'Tools for Change' *ie materials for:* • classrooms • assessment • professional development • persuasion • feedback – all aspects • ....

Note the crucial difference between T_1 level, which is about teaching possibilities, usually explored by a member of the research team, and T_2, which is about what can be achieved in practice by typical teachers with available levels of support. Currently, nearly all research is at L and T_1 levels. A better balance across the levels is needed, if research and practice are to benefit from each other as they could.

T_2 and C -level research needs larger research teams and longer time-scales, in harmony with the need to combine importance, generality and trustworthiness.

The status and roles of "theory"

Finally, some related comments on "theory", which is seen as the key mark of quality in educational research. I am strongly in favour of theory. (Indeed, in my other life, I am a theoretical physicist) However, in assessing its role, it is crucial to be clear as to how strong the theory is. From a practical point of view, the key question is:

How far is this theory an adequate basis for design?

Again it useful to look across fields. In Aeronautical Engineering. for example, the theory is strong; those who know the theory can design an aeroplane at a computer, build it, and it will fly, and fly efficiently. (They still flight test it extensively and exhaustively) In Medicine, theory is moderately weak, but getting stronger. Despite all that is known about physiology and pharmacology, much development is not theory-driven. The development of new drugs, for example, is still mainly done by testing the effects of very large numbers of naturally occurring substances; they are chosen intelligently, based on analogy with known drugs, but the effects are not predictable and the search is wide. However, as fundamental work on DNA has advanced, and with it the theoretical understanding of biological processes, designer drugs with much more theoretical input have begun to be developed. This process will continue – indeed there is now work, for example, on cancer drugs tailored to individual tumours.

Education is a long way behind medicine (100 years?), let alone engineering (350 years?), in the range and reliability of its theories. By overestimating their strength damage has been done to children, for example by designing curricula based largely on behaviourist theories. The current dominance of constructivism is similarly inadequate, though less dangerous (because its incompleteness is more obvious, since it seems impossible to design a curriculum built only from constructivist principles and a list of skills) It is not that behaviourism or constructivism are wrong; indeed, they are both right in their core ideas but they are incomplete and an inadequate basis for design. Physicists would call them 'effects'. The harm comes from overestimating their power.

Let me illustrate this with an example from meteorology. "Air flows from regions of high pressure to regions of low pressure" sounds and is good physics. It implies that air will come out of a popped balloon or a pump. It also implies *the effect that winds should blow perpendicular to the isobars*, the contour lines of pressure on a weather map, just as water flows downhill, perpendicular to the contour lines of a slope. However, a look at a good weather map shows that, in England, *the winds are closer to parallel to the isobars*. That is because *there is another effect*, the Coriolis Effect. It is due to the rotation of the earth which 'twists' the winds in a subtle way, clockwise around low pressure regions. (Like water down the plughole, they go round the other way in the Southern Hemisphere) In education there are many such effects operating. We have identified some of them (behaviourism and constructivism are two) but, as in economics, it is impossible to predict just how they will balance out in a given situation. Thus *design skill and empirical development are essential*, with *theoretical input providing useful heuristic guidance*. The essential point is that **the design details matter** – they have important effects on outcomes and are not determined by theory.

Empirical development is essential, usually *through several cycles* ("Fail Fast, Fail Often"), in *realistic circumstances*, mostly with *small samples* with *rich, detailed feedback* through observation and analysis . But such **care** is not enough; quality also needs **flair –**

outstanding design skill, which is rare and develops slowly. This is an **engineering research approach** to educational design. If effective implementation of the research presented at this conference is to happen, this approach will be needed, across the levels L to C.

A final comment. Good theories come in all scales of generality and range. Aeronautical Engineering uses Newton's Laws of Motion, a *general theory* which works well at all scales between the molecule (where Quantum Mechanics is needed) and the galaxy (where General relativity influences the behaviour significantly). It also uses *phenomenological theory*, such as the elastic properties of aluminium alloys – this provides essential *reliable input to design of limited but known range*.

There are two points here. First, theory in education of the latter type[1] is less sexy than general theory but currently more valuable in guiding design (and easier to establish). Fortunately, in recent years there has been a growing emphasis on research that produces such contributions though the desirable combination of range and trrustworthiness constraints of time, resources and research tradition. (It is probably no accident that much of this useful theory has come in work on computer-supported learning, where the higher development costs make these constraints less onerous)

Testing the above

This paper, as promised, is firmly in the humanities tradition. It is a set of critical comments and suggestions – assertions, based on experience and with some supportive examples. What kind of evidence will be needed to test these assertions, to answer some of the anticipated questions, and thus increasingly to justify a claim that it offers good engineering principles. There is space here for a few brief illustrations. Some things need, and justify, a lot of work; for others it is fairly simple and brutal, eg:.

How important is "good engineering"? For example: take some teaching materials, as used in a few exploratory insight research studies; give them to some typical teachers, who agree to use them for a short period with a regular class; observe the class and the student work; listen to the protests of the authors that these materials are not yet ready for that; point proven, or not.

How far do the details matters? This is a central question that is more interesting to explore. It seems worth doing so at different grain sizes, for example:

For a lesson, take an outstanding introductory lesson on a topic, and compare with variants, eg:; make a precis of its content and give it to three designers of varying experience (including the original author, who is asked to design a basic-if-boring lesson) Develop each draft version in classrooms. Compare in detail the student responses to the four developed versions, and the three

[1] A small example: We showed that classroom situations that move students into 'teacher roles' (Manager, Explainer, Task-setter) led naturally to their exhibiting and developing higher-level problem solving and metacognitive skills. This small piece of phenomenological theory has proved an invaluable heuristic in guiding design. (We did not establish trustworthy boundaries for this assertion, beyond the actual study – an important weakness)

drafts. This, replicated over different lessons, investigates the effects of both design flair and care in development. (If the details do matter, insight researchers may be led to consider using treatments built around well-engineered materials that can be transferred directly to typical classrooms – ie the integration of research styles suggested above)

For a curriculum, such as the complete teaching materials for a year's work, do a comparative investigation, in depth, of the patterns of classroom activities, student learning, performance and attitudes (of teachers and students) for several curricula designed on a common set of principles. (The NSF-funded US Middle School curricula seem good candidates. The UK National Numeracy materials exist in only one version, but could be usefully compared, in depth beyond the current narrow tests, with some elementary school standards)

This latter grain size needs big projects but would be of immense value in informing current purchasing and future planning. The few-lesson-scale investigations are tractable in small scale studies.

How effective are theoretical heuristics? Clearly, each heuristic has to be tackled individually, in various ways involving their use by other designers across a range of topics, teachers, students and other variables. For example, 'equivalent' lessons, with and without the use of the heuristic in question, should be designed by the same designer and compared in depth. Designers may well be motivated in such work, to find out what will improve their own design skills.

More examples to come plus

References