Chapter 9

From design research to large-scale impact

Engineering research in education

Hugh Burkhardt

Introduction

Previous chapters have described the current state of design research and its future prospects, setting out its contribution in bringing research in education closer to what actually happens in classrooms. Indeed, the last 30 years have been remarkable for the shift from the traditional combination of critical commentary on the one hand and laboratory experiments on the other, towards the empirical study of teachers and children in real classrooms. The growth of cognitive science and its application to research in more realistic learning environments has contributed much to this. The establishment of design research over the last decade represents the next step in this sequence. However, there is more to be done before teaching and learning *in the majority of classrooms* can possibly move to being research-based. How we may get there, and the progress so far, is the theme of this chapter.

First, we need an established research-based methodology for taking the design research approach forward to produce processes and tools that work well in practice with teachers and students who are typical of the target groups. I shall argue that this methodology is already in place, and describe how it works. This research approach is characteristic of engineering disciplines, with *new or better products and processes* as the primary outcomes, so we call it the *engineering research* approach.[1] Direct *impact* on practice is the main criterion of quality, though engineering research also delivers new *insights* – and journal articles. Second, we need reliable models of the process of educational change. These we do not yet have. I will outline progress that has been made, giving reasonable hope that a research-based developmental approach can succeed here too.

In what follows, we first begin with a comparison of the traditional *craft-based* approach to the improvement of professional practice with the *research-based* approach, going on to outline the key elements in the latter and how various groups contribute to it. The section that follows discusses the research infrastructure in education, the contributions of different research traditions to it, and the increased proportion of engineering research needed for the development of new and improved products and processes – and, through this, for impact on policy and practice. The next section describes the process of design, development and evaluation that is characteristic of this engineering research approach.

In the next two sections, we then discuss how one can build the skill base needed for such a program and analyse the implications for governments, industry and academia. The penultimate section sketches an answer to the question *How much would good engineering cost?*, while the final section summarizes the implications of all this for policy and for the design and development community. This wide-ranging agenda inevitably limits the detail in which the analysis can be justified, or exemplified; the still-too-limited evidence is thus mainly listed in the references.

R↔P: Research-based improvement of professional practice

Educational research still has much less impact on policy and practice than we would wish. If politicians have a problem in their education system, is their first move to call a research expert? Not often. Indeed, in most countries, there is no obvious link between changes in practice and any of the research of the tens of thousands of university researchers in education around the world. It is not like that in more research-based fields like medicine or engineering. Let us try to identify why.

Craft-based versus research-based approaches

I shall use two contrasting terms for improvement methodologies, *craft-based* and *research-based*. Like most dichotomies, it is an over-simplification but, I believe, useful.

In all professional fields, there is recognized *good practice*, embodying the established *craft skills* of the field. These are based on the collective experience of practitioners – they must always have a response to every situation that presents itself, whether they are teachers in a classroom, doctors in a surgery, or administrators running a system. Experienced professionals pass on good practices, and the skills involved, to new entrants in their training. This is what I mean by the *craft-based approach*. Historically, it was the approach of the craft *guilds* including, among many others, doctors and teachers. In this approach, innovation comes from a few people pushing the boundaries of good practice, trying something new and seeing if it works – for them. This sometimes involves the invention of new tools – instruments, teaching materials, etc. Others learn about it, and some try it; on the basis of this experience, they decide whether or not to adopt the innovation. If many take it up, it gradually becomes part of *good practice* – even then, it may be adopted by only a small minority of the profession.

All fields start with this approach, which has its strengths and limitations. It is inexpensive and anyone can take part. However, in judging an innovation, it lacks systematic evaluation of effectiveness in well-defined circumstances: for who, what, and when, does it *work*, and with what range of outcomes? Furthermore, it inevitably depends on the extrapolation of current experience in a *clinical* context; since any extrapolation is inherently unreliable, such exploration tends to be limited in scope.

Thus, the craft-based approach to innovation is limited in the range of possibilities it explores, and in the reliability of its conclusions. This has led to the search for more powerful and systematic, research-based approaches in many fields. Millennia ago, engineering took the lead in this. Starting in the late nineteenth century, medicine began to follow. Education set out on this path in the twentieth century. Other fields are also instructive. The clothing industry remains firmly craft-based, with changes driven by changes in fashion – hemlines go up and down, just as educational fashion swings back and forth between *basic skills* and *problem solving*; however, performance clothing for campers, climbers or astronauts is substantially research-based, and even has useful influence on some areas of fashion. For education, medicine seems a better model than the fashion industry.

Research-based approaches to improvement in large-scale educational practice are the theme of this chapter. Here it will suffice to say that its methods aim to:

- build on results from past research, as well as best practice;
- use research methods in a systematic process of exploring possibilities; then
- develop tools and processes for their use through creative design and successive refinement based on using research methods to get rich, detailed feedback in well-specified circumstances.

This is inevitably a slower and more costly approach than craft-based innovation. Slowly, over many years, research-based innovations gradually make an increasing contribution to the quality of practice in a field. Engineering is now largely research-based – bridges, aeroplanes and other products are designed from well-established theories and the known properties of materials. Over the last century, medicine has moved from being entirely craft-based to being substantially research-based. Fundamental discoveries, particularly in molecular biology, as well as a huge research effort have accelerated that process. Many treatments, in areas from the common cold to low-back pain, are still largely craft-based. Alternative/complementary medicine is almost entirely craft-based. However, the areas that have the firmer foundations of the research-based approach gradually expand. Education has only begun to move in this direction; nonetheless, progress has been made and is being made in both methodology and outcomes. What follows outlines some of it.

Key elements for research-based improvement

In a recent paper on 'Improving educational research' (Burkhardt and Schoenfeld (2003), referred to here as *IER*), we described the elements of R↔P mechanisms that are common to successful research-based fields of professional practice such as medicine and the design and engineering of consumer electronics. They all have robust mechanisms for taking ideas from laboratory scale to widely used practice. Such mechanisms typically involve multiple inputs from established research, the imaginative design of prototypes, refinement on the basis of feedback from systematic development, and marketing mechanisms that rely in part on respected third party in-depth evaluations. These lab-to-engineering-to-marketing linkages typically involve the academic community and a strong research-active industry (for example, the drug companies, Bell Labs, Xerox PARC, IBM and Google).

The following elements are all important in achieving effective and robust products:

- A body of reliable research, with a reasonably stable theoretical base, a minimum of faddishness and *a clear view of the reliable range of each aspect of the theory*. This in turn requires norms for research methods and reporting that are rigorous and consistent, resulting in a set of insights and/or prototype tools on which designers can rely. The goal, achieved in other fields, is *cumulativity* – a growing core of results, developed through studies that build on previous work, which are accepted by both the research community and the public as reliable and non-controversial within a well-defined range of circumstances. Such a theory base allows for a clear focus on important issues and provides sound (though still limited) guidance for the design of improved solutions to important problems.
- Stable design teams of adequate size to grapple with large tasks over the relatively long time scales required for sound work of major importance in both research and development. Informed by the research base and good practice, they add another crucial ingredient – design skill, even brilliance.
- Systematic iterative development that takes the tools through successive rounds of trials and revision in increasingly realistic circumstances of use and users.
- Independent comparative in-depth evaluation provides validation (or not): Do they work as claimed, in the range of circumstances claimed? This provides the basis for the following point.
- Individual and group accountability for ideas and products, so that reputations are built on a track record of evidence.

It should be clear that this approach requires ongoing development programs on realistic time scales, funded by clients who understand the process. To do it well needs substantial teams; it cannot be done by individual researchers in a few years – the normal circumstances of university research in education.

Around the world there are some well-established high-quality engineering research groups – EDC, TERC and COMAP are notable U.S. examples; however, these and most other groups work from project to project, with no continuity of funding or, consequently, of work. Only the Freudenthal Institute has had a substantial team (now about 70 people), supported by continuing funding from the Netherlands Government over many decades;[2] the quality of their work on basic research, design and engineering is universally recognized – and reflected in the performance of the Netherlands in international comparison tests in mathematics. It represents the best current exemplar for governments elsewhere to study.

Key contributors – roles and barriers

In this approach, who are the key players, what are their roles, and current barriers to the fulfilment of these roles?

Client funders, as well as providing the money that supports the work, should be partners in goal and product definition through a continuing process of negotiation that reconciles their goals and design constraints in the best solution that can be devised; the main barriers to this are often an unquestioned acceptance of the traditional craft-based approach, left undisturbed by the lack of comparative in-depth evaluation of the effectiveness of products, or an insistence on ‘simple solutions’ that will not, in practice, meet their goals.

Project leaders guide the strategic planning and ongoing direction of their team’s work. This involves negotiation with funders, and process management at all levels from design to marketing. The main barriers to project leaders’ development are the lack of continuity produced by one-off project funding, and the consequent absence of a career path for engineering researchers – that is, systems for training and apprenticeship, appointment and promotion, and recognition.

A *designer/developer* provides excellence in design, and refinement through feedback from trials; the main barriers are, again, the lack of any career path for designers or the in-depth evaluation that would enable the recognition of excellence in design.

Insight-focused researchers build the reliable research base, and carry out the comparative in-depth evaluation, both formative and summative, of products, which is so important; the main barrier is the academic system, which undervalues such project-focused work, and also contributes to the other barriers above.

Each group needs to play these vital roles. In the section ‘Changing behaviour in academia, industry, governments' (p. 142) we look at the changes that are needed in their current working environments to make this possible.

The research infrastructure – insight versus impact as research goals

At a fundamental level, the relative impotence of research in education arises from the interaction of different research traditions and styles, characteristic respectively of the humanities, sciences, engineering and the arts. For this analysis, we need to go beyond the familiar controversies and *paradigm wars* in education; well-organized fields recognize that strength in research requires a wide range of approaches, tailored to the problems at hand. Let us take a broader view, looking across fields at the four characteristic research styles and asking how each contributes to education. For this, it is useful to have a definition of research (HEFC 1999), designed to cover all fields:[3]

> 'Research' 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.
>
> (RAE 2001)

If you then look for a fundamental measure of *quality* in research across all fields, it is difficult to go beyond *impressing key people in your field* – but the balance of qualities that achieves this varies from field to field. What balance would be most beneficial for education, and how well is it reflected in current criteria for excellence in research? Let us look at each style in turn, the nature of the activities, the forms of output and, in context of education, the potential impact on students' learning in typical classrooms.

The humanities *approach*

This is the oldest research tradition, summarized in the aforementioned RAE (2001) exercise as '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.' Empirical testing of the assertions made is *not* involved. The key product is *critical* *commentary*, usually published in single-author books, journal papers or, indeed, journalism.

There is a lot of this in education, partly because anyone can play at making assertions, expert or not; indeed, there is no popular acceptance of expertise. The ideas and analysis in the best work of this kind, based on the authors' observation and reflections on their experience, are 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 worked in practice, the lack of empirical support is a major weakness. How can you distinguish reliable comment from plausible speculation? This has led to a search for *evidence-based education*.

The science *approach*

This approach to research is also focused on the development of better *insights* and improved understanding of *how the world works* through the analysis of phenomena and the building of models which explain them, but in this case it includes the *empirical testing of those models*. This last is the essential difference from the humanities approach – the assertions made, now called hypotheses or models, depend on rigorous empirical testing for credibility. The main products are: *assertions with evidence-based arguments in support,* including *evidence-based responses to key questions*. The evidence must be empirical, and presented in a form that could be replicated. The products are conference talks and journal papers.

Schoenfeld (2002) 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 three-dimensional space. Importance, a key variable, could usefully distinguish *insight* from *impact*. Figure 9.1 focuses on the other two variables, G and T, say. A typical research study looks carefully at a particular situation, perhaps a specific treatment and student responses to it. The results are high on T, low on G – zone A in the figure. Typically, the conclusions section of the paper goes on to discuss the *implications* of the study, often much more wide-ranging but with little evidence to support the generalizations made, which are essentially speculative (in the humanities tradition) – shown as X, Y and Z.

A lot of research really provides evidence *on treatments*, *not on the principles* the authors claim to study; to probe the latter, one needs evidence on

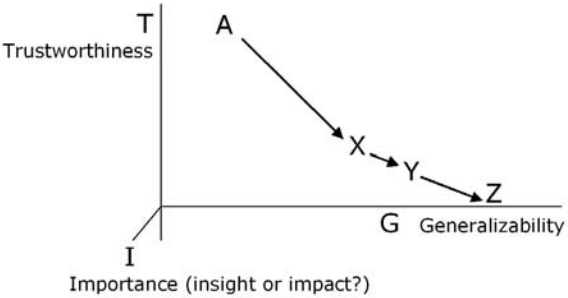


Figure 9.1 Graphical representation of typical study outputs (based on Schoenfeld 2002)

generalizability – one must check stability across a range of variables (student, teacher, designer and topic). Only substantial studies, or sometimes meta-analyses, can meet this need. The 'Diagnostic teaching' work of Alan Bell, Malcolm Swan and others illustrate this. The approach is based on leading students whose conceptual understanding is not yet robust into making errors, then helping them understand and debug them through discussion – see, for example, Bell (1993). Their first study compared a diagnostic teaching treatment to a standard 'positive only' teaching approach. It showed similar learning gains through the teaching period (pre- to post-test) but without the fall away over the following 6 months that the comparison group showed. This 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 statements about diagnostic teaching as an approach. Few studies persist in this way. Even then, further questions will remain – in this case, as so often, about how well *typical* teachers in realistic circumstances of support will handle diagnostic teaching. Work on this continues.

My general point is that insight-focused research with adequate evidence on its range of validity needs time and teams beyond the scale of an individual Ph.D. or research grant. Other academic subjects, from molecular biology to high-energy physics, 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 the serious attention of designers and policymakers.

There is the further issue of grain size. Studies of student learning in tightly controlled laboratory conditions are too artificial to use directly in guiding design for the classroom. At the other extreme, studies of student performance on complete programs tell you little about the classroom causes. Design research addresses the key problems directly.

The science approach is now predominant in research in science and mathematics education; it is not yet influential in policy formation. Such research provides insights, identifies problems and suggests possible ways forward. Design research takes this one step further into more realistic learning environments; it does not itself generate *practical solutions for situations that are typical of the system* – that also needs good engineering.

The engineering *approach*

This approach aims to go beyond improved insights to direct practical *impact* – helping the world *to work better* by, not only understanding how it works, but developing *robust* solutions to recognized practical problems. It builds on science insights, insofar as they are available, but goes beyond them. Within the broad RAE (2001) definition mentioned previously 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 the empirical testing of products and processes, both formatively in the development process and in evaluation. 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, and only when, it includes these elements, is development engineering research.

This approach is still uncommon in education – though there are many good examples, it is not the way most challenges are tackled.[4] In the academic education community such work is often undervalued – in many places only insight-focused research in the science tradition is regarded as true research currency. In this environment, it is not surprising that most design research stresses the new insights it provides rather than the products and processes it has developed, even though these could be valuable if developed further. The effects of the current low academic status of educational engineering include:

- lower standards of materials and processes, since the imaginative design and rigorous development that good engineering requires are not widely demanded;
- lower practical impact of important results from insight-focused research, since designers feel less need to know or use this research; and
- pressure on academics in universities to produce insight research papers, rather than use engineering research methods. (These could also be used to improve their own practice, both in effectiveness and transferability to others.)

All this leaves a hiatus between insight-focused research and improved classroom practice, which is unfortunate. Society's priorities for education are mainly practical – that young people should learn to function in life as effectively as possible, including the personal satisfaction and growth that good education provides. The failure of educational research to deliver in practical terms is reflected in the low levels of financial support for it.

The arts *approach*

A change in the criteria of excellence in fine arts is noteworthy. Fifty years ago, only critics and historians of art or music would be appointed to senior academic positions; now active artists, painters and composers, are appointed – as are designers and innovative practitioners in engineering and medicine. This may be seen as related to the humanities approach rather as engineering is to the science approach. It reminds us that design is more than the routine application of a set of scientific principles. (Indeed, the finest engineering, from the Porsche to the iPod, has a strong aesthetic aspect.) It enriches education and could do more.

Integrating the traditions

In summary, let me stress that this is far from a plea for the abandonment of insight-focused science research in education, or even the critical commentary of the humanities tradition; these are essential, *but not nearly enough*. Rather, it is an argument about balance – that there should be much more impact-focused engineering research and that it should receive comparable recognition and reward. *The different styles can and should be complementary and mutually supportive*. But more engineering research is essential if impact is to be a research priority of the field. As we deliver impact, our work will become more useful to practice, more influential on policy – and, as other fields have shown, much better funded.

The status and roles of theory

Finally, some related comments on *theory*, seen as the key mark of quality in educational research as in most fields. 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 current theory an adequate basis for design?*

A strong theory provides an explanation of what is behind an array of observations. It is complete enough to model the behaviour it explains, to predict outputs from conditions and inputs. In fields like aeronautical engineering, the theory is strong; the model is complete enough to handle nearly all the relevant variables so that 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, so that trial design and testing is more central. 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. Looking across fields, it seems that the power of theory and the engineering research approach develop in parallel.

Education is a long way behind medicine, let alone engineering, 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. It is not that behaviourism, or constructivism, is wrong; indeed, they are both right in their core ideas but they are *incomplete* and, on their own, make 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 like and is good physics. It implies that air will come out of a popped balloon or a pump. It also implies that winds should blow perpendicular to the isobars, the contour lines of equal 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 in England shows that 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, anticlockwise around low pressure regions (in the northern hemisphere). There are many such effects operating in education 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 guided, not determined, by theory.[5]

This paper is about how to achieve educational goals. I will not explicitly discuss the goals themselves. This is only partly because the subject is huge and disputatious. I believe that, in mathematics and science at least, many of the apparently goal-focused disputes are in fact based on strongly held beliefs about how to achieve them, some of which fly in the face of the research evidence. For example, those who think mathematics should focus on procedural ‘basic skills’ in arithmetic and algebra usually want students to be able to solve real-world problems with mathematics, but think ‘they must have a firm foundation of skills first.’ It is true that some with a strong faith-based world view believe that schools should not encourage students to question authority – an essential aspect of problem solving and investigation; however, even here, the greater challenge is to equip teachers to handle investigative work in their classroom. There is, therefore, an implicit assumption here that the educational goals we address are those that research on learning and teaching suggest are essential – a challenging enough set.

Systematic design, development and evaluation

From this analysis of framework and infrastructure, it is time to move on to look in more detail at engineering research itself – the methodology that enables it to produce high-quality tools, with processes for their effective use. The approach is based on a fusion of the elements already discussed:

- research input from earlier research and development worldwide;
- design skill, led by designers who have produced exceptional materials;
- co-development with members of the target communities;
- rich, detailed feedback from successive rounds of developmental trials to guide revision of the materials, so that intentions and outcomes converge; and
- a well-defined locus of 'design control', so that wide consultation can be combined with design coherence.

Typically, there are three stages: design, systematic development and evaluation. I discuss each in turn, with brief exemplification from one project. Research of various kinds plays several roles: input to design from earlier research; research methods for the development process; in-depth research for evaluation, to inform users on product selection and the design community on future development.

The example I shall use is from the (still ongoing) development of support for *mathematical literacy*, now called *functional mathematics* in England. Our work began in the 1980s with a project called *Numeracy Through Problem Solving* (NTPS; Shell Centre 1987–89). U.K. Government interest has recently revived (Tomlinson Report 2004), partly through the emergence of PISA (OECD 2003) and (yet again) of employers' concern at the non-functionality of their employees' mathematics.

Design

The importance of sound design principles, based on the best insight research, has long been clear. They are necessary but not sufficient for the production of excellent tools for practitioners, essentially because of our current far-from-complete understanding of learning and teaching. The other key factor is excellence in design; it makes the difference between an acceptable-but-mediocre product and one that is outstanding, empowering the users and lifting their spirits.

Excellent design is balanced across the educational goals, covering both functional effectiveness and aesthetic attractiveness. (Porsches are the wonderful cars they are because they're classy *and* superbly engineered. There are sleek-looking cars one wouldn't want to own, for very long at least.) As in every field, design that is aimed at superficial aspects of the product (textbooks in 4-color editions, with lots of sidebars with historical and other trivia, pictures that have no connection with the topic, etc.) at the expense of effectiveness (in promoting learning) is poor educational design. (Focused on increasing sales rather than increasing student understanding, they may be good *business* design.)

In the humanities and the arts, teachers can build their lessons around outstanding works – of literature or music, for example, in their infinite variety. Because the learning focus in mathematics and science is on concepts and theories rather than their diverse realizations, excellence has to be designed into the teaching materials. How to achieve such excellence is less well understood, or researched – partly because its importance is not widely recognized in education.

Design skill can be developed but it is partly innate, born not made. It grows with experience over many years. Outstanding designers seem to work in different ways, some being mainly driven by theoretical ideas, others by previous exemplars, or by inspiration from the world around them. All have the ability to integrate multiple inputs to their imagination. All have deep understanding of the craft skills of the environment they design for – mathematics or science teaching, for example. Quality seems to lie in combining specific learning foci with a rich complexity of connections to other ideas, integrated in a natural-seeming way that feels easy in use. Not so different from literature and music.

Design excellence is recognizable; people tend to agree on which products show design flair. At the present crude level of understanding of design, the best advice to project leaders and institutions is heuristic – look for outstanding designers, and give them an environment in which they flourish and develop. Above all, if you want outstanding products, don't overdirect designers with detailed design objectives and constraints; balancing and fine-tuning can be done later by other, analytic minds. Keep each design team small so that communication is through day-by-day conversation, rather than management structures. (The extraordinary research creativity of Xerox PARC was achieved with a maximum of ten people per research team.)

The stages of the design phase are typically:

- outlining an agreement with the client on the broad goals and structure of the product;

- generating design ideas within the design group, in consultation with experts and outstanding practitioners; and
- drafting materials, which are tried out in the target arena (for example, classrooms) by the lead designer and others in the design group, then revised, producing the all-important ‘alpha version’.

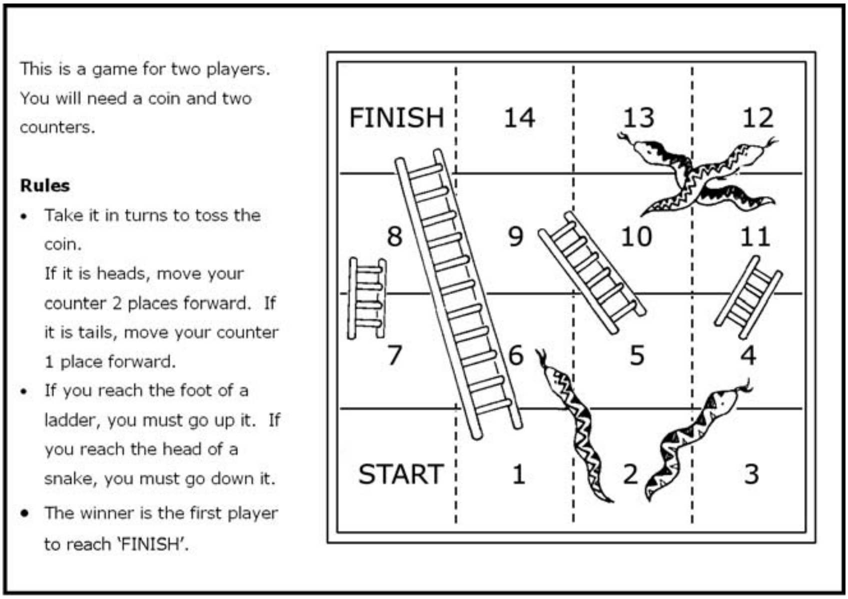
In practice, as always with creative processes, there is cycling among these stages.

Design control is a concept we at the Shell Centre have found to be important to the progress of any project. The principle is that one person, after appropriate consultation with the team, takes all the design decisions on the aspect they control. This has two major advantages. It retains design coherence, which improves quality, and it avoids extended debates in the search for consensus, which saves time and energy. If a consensus is clear in discussion, the designer is expected to follow it – or have a very good explanation for doing something else. (Everyone is expected to take very seriously the empirical feedback from the trials.)

NTPS moved through the design phase in the following way. In the mid-1980s, there was general discontent with the national system of assessment in mathematics at age sixteen, stimulated by the Cockcroft Report (1979). There was an opportunity to try new examinations, linked to a recognition that the academic remoteness of mathematics was not ideal or essential – that functional mathematics should be tried. I proposed to the Joint Matriculation Board that the Shell Centre team should develop a new assessment with them and, because it breaks new ground, teaching materials to enable teachers to prepare for it. We agreed on five to ten 3-week modules, each with its own assessment, both during the module and afterwards.

We assembled a group of six innovative teachers and, with a few outside experts on mathematics education, held a series of brainstorming sessions on topics from the outside world that might make good modules. About thirty topics were considered; individuals or pairs drafted rough few-page outlines for each topic. After much discussion, we settled on ten topics for further exploratory development, with a fairly clear view of their order. (In the end, there were five.)

Design control was clarified. Malcolm Swan would lead the design, particularly of the student materials and assessment tasks. John Gillespie organized the trials and relations with schools, and led the design of one module. Barbara Binns wrote notes for teachers, and managed the development process, including links with the examination board. I led on strategic issues (for example, an initial challenge was to ensure that the design remained focused on the *unfamiliar* goal, functional mathematics in the service of real problem solving, rather than reverting to ‘just mathematics’), on the overall structure of the product, and how we would get there. Everyone contributed ideas and suggestions on all aspects of the work.



This is a game for two players.
You will need a coin and two counters.

Rules

- Take it in turns to toss the coin.
  If it is heads, move your counter 2 places forward. If it is tails, move your counter 1 place forward.
- If you reach the foot of a ladder, you must go up it. If you reach the head of a snake, you must go down it.
- The winner is the first player to reach 'FINISH'.

Figure 9.2 Snakes and Ladders assessment task

In understanding the challenge of this kind of problem, we decided to break each module into four stages, characteristic of good problem solving: understanding the problem situation; brainstorming; detailed design and planning; implementation and evaluation. A key challenge in all investigative work is to sustain students' autonomy as problem solvers, without their losing their way or being discouraged. We decided that students should work in groups of three or four, guided by a student booklet. Individual assessment at the end of each stage would monitor the understanding of each student. Among many design details, the booklet gave strategic guidance on what to do in each stage, with delayed checklists to ensure that nothing essential had been overlooked.

'Design a board game' was chosen as the first module to develop. Understanding of what this involves was achieved by creating a series of amusingly bad board games for the students to play, critique, and improve. (The students were delighted that these 'wrong answers' came from the examination board – an unexpected bonus.) The *Snakes and Ladders* assessment task in Figure 9.2 exemplifies this.

They now understood that a game needs a board and a set of rules, it should be fair – and should end in a reasonable time! Each group then enjoyed exploring a range of ideas. The design and construction of their board, and testing the game followed. (In one enterprising trial school, this became a joint project with the art department.) Evaluation was accomplished by each group playing the other groups' games, commenting on them, and voting for a favourite. Notes for teachers had been built up by the team through this tryout process. These and the student books were revised and assembled into first draft form, ready for trials.

The 'final examinations' were designed rather later. There were two papers for each module, called Standard and Extension Levels, which assessed the student's ability to transfer what they had learnt, to less- and more-remote problem situations respectively. These were externally scored by the board. Basic level was awarded on the basis of the assessment tasks embedded in the teaching materials. Assessing the group's products was seen as a step too far.

To summarize, it is the integration of research-based design principles and excellence in design with appropriate educational goals that produces really exceptional educational products.

Systematic iterative development

The design process produces draft materials. The team has some evidence on the response of students, albeit with atypical teachers (the authors), but none on how well the materials transfer, helping other teachers create comparable learning experiences in their classrooms. It is systematic development that turns drafts into robust and effective products. It involves successive rounds of trials, with rich and detailed feedback, in increasingly realistic circumstances.

The feedback at each stage guides the revision of the materials by the design team. Feedback can take many forms; the criterion for choosing what information to collect is its usefulness for that purpose. This also depends on presenting it in a form that the designers can readily absorb – too much indigestible information is as useless as too little; equally, it depends on the designers' willingness to learn from feedback, and having the skills to infer appropriate design changes from it. Cost-effectiveness then implies different balances of feedback at each stage. In the development of teaching materials, these typically include alpha trials, followed by discussion, and then beta trials, followed by discussion, until the final version is ready for publication.

Alpha trials were held in a handful of classrooms (normally five to ten), some with robust teachers who can handle anything and others more typical of the target group. This small number is enough to allow observers to distinguish those things that are generic, found in most of the classrooms, from those that are idiosyncratic. The priority at this stage is the quality of feedback from each classroom, including:

- structured observation reports by a team of observers, covering in detail every lesson of each teacher;
- samples of student work, for analysis by the team; and

- informal-but-structured interviews with teachers and students on their overall response to the lesson, and on the details of the lesson materials, line by line.

The process of communicating what has been found to the designers is important, and difficult to optimize. We like to have meetings, in which the observers share their information with the lead designer in two stages presenting: first, an analytic picture of each teacher in the trials, working without and with the new materials; then, a line-by-line discussion of the materials, bringing out what happened in each of the classrooms, noting where the materials did not communicate effectively to teacher or students and how the intended activities worked out. The discussion in these sessions is primarily about clarifying the meaning of the data, but suggestions for revision also flow. The role of the lead designer in this process is that of listener and questioner, absorbing the information and suggestions, and integrating them into decisions on revision.

Revision by the lead designer follows this discussion, producing the ‘beta version’. The priorities are different now, focused on the realization of the lessons in typical classrooms. A larger sample (twenty to fifty) is needed. It should be roughly representative of the target groups. (We have usually obtained stratified, reasonably random samples by invitation – “You have been chosen ....” has good acceptance rates, particularly when the materials can be related to high-stakes assessment.) Within given team resources, a larger sample means more limited feedback from each classroom, largely confined to written material from samples of students. Observation of the beta version in use in another small group of classrooms is an important complement to this.

Revision by the lead designer again follows, producing the final version for publication. The development of NTPS worked very much in this way. Numerous improvements were made as the result of the feedback from the alpha trials, some removing ‘bugs’ in the activities themselves, or in the teacher’s misunderstanding of the guidance, others incorporating good ideas that emerged from individual classrooms. The beta trials were mostly checking and validating what we had learnt with the larger sample – they produced many small changes. The evaluative feedback we received provided a substantial basis for a summative view of the outcomes, positive in terms of student achievement and (vividly) of attitude to mathematics. A notable feature was the narrowing gap between previously high- and low-performing students – an important equity goal that is notoriously hard to achieve. It seemed to arise largely from discomfort with non-routine tasks for some, and improved motivation for others. Because of pressures and priorities, none of this data was collected in a sufficiently structured way for a research journal – a defect, common in such work, that one would like to have the time and resources to overcome.

This is not the end of the process. Feedback 'from the field' will guide future developments. Both informal comments from users and more structured research will produce insights on which to build. Changing circumstances may lead to further development of the product. NTPS, with its own examination, was sidelined by the introduction of the GCSE as a universal examination at age sixteen. To continue to serve the schools that had become enthusiastic about this functional approach to mathematics, we and the examination board developed a GCSE built around it. As so often, fitting into this new framework led to compromises and some distortion of the approach, with more emphasis on imitative exercises rather than mathematics in use.

Such an engineering research methodology is common in many fields for the development of tools and processes so as to ensure that they work well for their intended users and purposes. It is still often neglected in education for the craft-based approach, which may be summarized as: write draft materials from your own experience; circulate to an expert group; discuss at meetings; revise; publish. This is quicker and cheaper, but does not allow substantial innovations that work effectively for the whole target community of users.

Weak design and development can produce costly flaws. (For example, the unintended consequences of pressure for 'simple tests' in mathematics include a destructive fragmentation of learning as teachers 'teach to the test'.) And it is well known in engineering that the later a flaw is detected, the more it costs to fix – more by orders of magnitude!

Comparative in-depth evaluation

This third key element in the engineering approach is also the least developed.[6] It is nonetheless critically important for *policy makers and practitioners*, guiding choices of materials and approaches, and for *design teams*, informing product improvement and future developments. For the first of these, evaluation needs to be seen to be independent; for both it needs to look in depth at: (1) widely available treatments, competing in the same area; (2) all important variables: types of user, styles of use, and levels of support (professional development etc.); (3) outcome measures that cover the full range of design intentions, including classroom activities as well as student performance; and (4) alternative products, their approaches and detailed engineering.

The scale implied in this specification explains why, as far as I know, there has been no such study, anywhere in the world, although the research skills it needs are in the mainstream of current educational research.

In a back-of-the-envelope look at what such a study in the United States of America might entail, I estimate:

- Time scale: year 1 preparation and recruiting; year 2 piloting, schools start curriculum; year 3–6 data capture; year 4–7 analysis and publication; year 5–7 curriculum revision; then loop to year 3 with some new materials;

- Grade range: three middle grades, ages 11–14 (others a year or two behind);
- Curricula: nine diverse published curricula; two or three focal units per year in each curriculum;
- School systems: ten nationwide, diverse, subject to agreement to go with the methodology;
- Schools and classrooms: two classrooms per grade in ten schools per system, assigned in pairs to five of the curricula "by invitation," with levels of professional development and other support that the school system agrees to reproduce on large-scale implementation; and
- Data: Pre-, post- and delayed-test scores on a broad range of assessment instruments; ongoing samples of student work; classroom observation of ten lessons per year in each class, with post-interviews; pre-, post- and delayed-questionnaires on beliefs, style and attitudes.

A rough estimate of the cost of such an exercise focuses on lesson observation, the most expensive component. Assuming one observer–researcher for eight classrooms (that is, eighty observations per year, plus all the other work), this implies for each grade range: 10 school systems × 20 classrooms × 3 grades/8 = 75 researchers @ $100K per year ~$10 million a year for the 3 main years (three to six) of the study ~$30 million including leadership, support and overhead. With five universities involved, each would need a team of about fifteen people covering the necessary range of research, development and system skills. This is, indeed, 'big education' – but likely to be cost-effective (see later section on the costs of good engineering, p. 146).

Functional mathematics will need this kind of evaluation in due time, bearing in mind that a typical time from agreeing goals to stable curriculum implementation is approximately ten years. Meanwhile, it represents a research and development challenge.

Systematic development of models for system change

In the introduction, I noted the need for reliable, research-based models of the overall process of educational change – approaches that are validated, not merely by *post hoc* analyses but by evidential warrants for robustness that policy makers can rely on. This is clearly a challenging design and development problem. Nowhere has it been solved for the kind of educational changes that research has shown to be essential for high-quality learning, at least in mathematics and science. Some progress has been made, giving reasonable hope that a similar developmental approach can succeed.

Here the system of study is much more complex (even) than the classroom or the professional development of teachers, involving a much broader range of key players – students, teachers, principals, professional leadership, system administrators, politicians and the public. As with any planned change, all of the key groups must move in the way intended, if the outcomes are to resemble the intentions.

There are now well-engineered exemplars of many of the key elements in such a change. Some are familiar tools – classroom teaching materials, and assessment that will encourage and reward aspects of performance that reflect the new goals, have long been recognized as essential support for large-scale change. In recent years, tools have been developed in some other areas that have previously been seen as inevitably craft-based. For example, materials to support specific kinds of 'live' professional development have been developed, and shown to enable less-experienced leaders to replace experts without substantial 'loss' to the participants. Given that there are a thousand mathematics teachers for every such expert, this is an important step forward in seeking large-scale improvement.

The focus has now moved beyond such specific areas to the change process itself. Until now, support for systems has been in the form of general advice and occasional 'technical assistance' by 'experts'. We have begun to develop a 'Toolkit for Change Agents' (see www.toolkitforchange.org), which aims to suggest successful *strategies* for responding to the common *challenges* that inevitably arise in every improvement program, and the *tools* that each employs. The entries in the toolkit are based on the successful experience of other 'change agents' who faced similar challenges. This work is still at an early stage but shows promise of helping with this core problem.

However, my purpose here is mainly to draw attention to the educational change process as an area that needs more than the critical commentary that has guided it so far. Because the system of study is larger than a student or a classroom, with more obviously important variables, the challenge to systematic research and development is greater. However, it is surely possible to provide those seeking to promote improvement with well-engineered tools that will increase their effectiveness, founded on research-based insights into the processes of change. This, too, will need large-scale projects.

Building the skill base for engineering research

The number of groups capable of high-quality engineering is now small and, as we have noted, they are far from secure. If the research-based approach is to become a substantial and effective part of large-scale improvement, the number of those who can do such work will need to grow. This will involve both finding and training people with specific skills and creating institutional structures that can handle such work as well as the teams of people with complementary skills that it requires. This takes time, typically a decade or so, fortunately matching the time any new approach will need to build public confidence in its value. We also need to bear in mind the changing balance of work.

Four levels of R&D – improving balance

I noted that a focus on improving practice will need a different balance of effort among the research styles presented earlier (insight versus impact), with more engineering research. A complementary perspective on balance is provided by looking at different 'levels' of research (R) and development (D), summarized in Table 9.1.

Note the crucial difference between ET, which is about teaching possibilities, usually explored by a member of the research team, and RT, which is about what can be achieved in practice by typical teachers with realistic levels of support. Note how the research foci, R in the third column, change across the levels. Currently, nearly all research is at L and ET levels. A better balance across the levels is needed, if research and practice are to benefit from each other as they could. The main contribution of design research has been to link the R and D elements in the third column – but, in most cases, only for the first two levels, L and ET. Both RT and SC research need larger

Table 9.1 Four levels of R&D

Level	*Variables*	*Typical research and development foci*
Learning (L)	Student Task	R: concepts, skills strategies, metacognition D: learning situations, probes, data capture
Exceptional Teacher (ET)	Instruction Student Task	R: teaching tactics + strategies, nature of student learning D: classroom materials for some teachers
Representative Teacher (RT)	Teacher Instruction Student Task	R: performance of representative teachers with realistic support; basic studies of teacher competencies D: classroom materials that 'work' for most teachers
System Change (SC)	System School Teacher Instruction Student Task	R: system change D: 'Tools for Change', i.e. materials for: classrooms, assessment, professional development, community relations

research teams and longer time scales, which is difficult to accommodate within typical academic structures, current in education.

What skills, and where will they come from?

In the first section (p. 122) I noted the key groups of contributors. Let us look at where they are needed, and where they will come from.

Insight-focused researchers with the necessary range of skills for their roles in the engineering approach already exist in large numbers in universities; the challenge, to create an academic climate that will encourage them to do such work, is discussed in the next section.

Designer–developers of high quality are rare,[9] partly because of the lack of any career path, from apprentice to expert professional, that encourages this activity; the development of this area, and the understanding of design skill in education, is still at an early stage. Progress in this area will be an important factor in the whole enterprise (Gardner and Shulman 2005). An International Society for Design and Development in Education (ISDDE; www.ISDDE.org) has recently been founded with the goals of:

- improving the design and development process
- building a design and development community
- increasing the impact of this on educational practice.

Project leaders are a similarly rare species, for much the same reasons – the multidimensional skills needed for this work are fairly well understood but, even with a supportive environment, it will take time to develop project leaders within the design and development community with experience in educational engineering.

Client funders with understanding of good engineering will appear as its potential is recognized – indeed, the scale of the funding of long-term coherent programs may be the best measure of progress; the ability of the engineering community to demonstrate this potential through funded projects will be crucial to justifying expansion.

All of these groups play vital roles. In the following section, we look at the changes that are needed in their current working environments to make progress possible.

Changing behaviour in academia, industry and governments

If these things were happening in major education systems, there would be no need for this paper. My colleagues and I could just concentrate on good engineering. Currently, moving from the present to the kind of approach outlined above will require change by all the key players in educational innovation. It is simplest to discuss them in reverse order.

Governments

Experience in other fields suggests that substantial government funding will flow into research in education, if and when policy makers and the public become convinced that the research community can deliver clear and important practical benefits to the system. (*IER* discusses this in some detail.) Medical research only received significant support from the early twentieth century on, as research-based benefits such as X-rays began to appear. Massive support followed the impact of penicillin and other antibiotics after 1945, perhaps helped by the drama of its discovery by Alexander Fleming. Physics too, particularly nuclear physics, was a fairly abstruse field until that time. (The annual budget in the 1930s of the world-leading Cavendish Laboratory under Rutherford was about £3,000 – peanuts by today's standards, even in real terms.) The role of physicists in World War II in the development of radar, operations research, nuclear weapons and many other things increased funding for pure, as well as applied, physics research by many orders of magnitude – a situation that continues to this day. (It has also continued to spin off practical benefits, including the founders of molecular biology, the Internet, and the World Wide Web.)

However, in most of these cases, government played a crucial pump-priming role, providing funding for 'proof of concept' studies while the practical benefits were still unproven. That will be necessary in education; however, it will only make sense to policy makers if credible structures are in place that give real promise of clear and direct practical benefits in the medium term. This will need a growing body of exemplar products of well-recognized effectiveness.

Industry

Here there is a different problem. In medicine and engineering, for example, there are industries that turn the prototypes of academic research into fully developed practical tools and processes. Pharmaceuticals and electronics are two obvious examples but the same is true across manufacturing industry. Firms have established links with academic researchers in their fields; they support pure research and the research-based development of prototypes. The firms then take these through the long and costly process of development into robust products.

No comparable industry exists in education. The publishing industry (the obvious candidate) turns prototypes (manuscripts) into products (books), but with minimal development – typically comments by a few 'experts' and a small-scale trial with teachers who, again, are simply asked to comment. Neither you nor the regulator would allow your children to be treated with a drug, or fly in a plane, that had been developed like this. It produces products that *work*, in some sense, but it is no way to break new ground with products that are really effective – both well designed and robust in use.[10]

Why is this so? The main reason is the continuing dominance of the craft-based approach, surviving largely because of the inadequate evaluation process. There is no systematic independent testing and reporting on the effectiveness of products. New teaching materials *are* regularly reviewed – but by an 'expert' who must deliver the review in a week or two, purely on the basis of inspection. The improved effectiveness produced by development often depends on quite subtle refinements; it only shows when representative samples of users (for example, typical teachers and students, working with the materials), are studied in depth. As we have noted, such studies take time and cost money. The situation is exacerbated because the *buyer* and the *user* of many products are different – for example, the school system buys the teaching materials that the teachers use. Marketing is, of course, aimed at the buyer. Given this situation, there is a greater need, and responsibility, for the academic community to do this kind of work – and for governments to fund it.[11] It goes almost without saying that there are no regulatory agencies on the lines of those that every country has for drugs and for aeroplanes.

Thus, currently there is no incentive for industry to invest in systematic development. Systematic evaluation, preferably before marketing, would change that; it would increase the cost of materials but not to a strategically significant extent (see the next section, p. 146).

Academia

If the situation is to improve, major changes will need to come in academia. Currently the *academic value system in education*, which controls academic appointments and promotions, is actively hostile to engineering research. As discussed in *IER*, it favours:

- new ideas *over* reliable research
- new results *over* replication and extension
- trustworthiness *over* generalizability
- small studies *over* major programs
- personal research *over* team research
- first author *over* team member
- disputation *over* consensus building
- journal papers *over* products and processes.

Schoenfeld (2002) describes most such studies as of 'limited generality but … (if properly done) … "here is something worth paying attention to"'. As I have noted, that is a totally inadequate basis for design. In all respects these values undermine research that would have clear impact on the improvement of teaching and learning.

A status pattern, where the pure is valued far more than the applied, is common but it is not general at any level of research. Many Nobel Prizes are for the design and development of devices – for example, only two people have won *two* Nobel Prizes in the same field: John Bardeen, the physicist, for the *transistor*, and for the *theory of superconductivity* and 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. With examples like these, education need not fear for its respectability in giving equal status to engineering research. These lie in 'Pasteur's Quadrant' (Stokes 1997) of work that contributes both practical benefits and new insights. However, one should not undervalue work in Edison's Quadrant, with its purely practical focus – contributions like the luminous filament light bulb are of inestimable social value. Note also that, in making this discovery, Edison investigated and catalogued the properties of hundreds of other candidate materials, adding to the body of phenomenological knowledge that is part of the theoretical underpinning of all engineering. In contrast, so much research in education lies in the quadrant that has no name – advancing neither theory nor practice.

Changing the culture in any established profession is notoriously difficult. What actions may help to bring this about in educational research? Leaders in the academic research community can make a major contribution by including direct impact on practice as a key criterion for judging research, complementing valid current criteria. One may (hopefully) envisage a future search committee at an academic institution that wants to hire a senior person in education, and is mindful of public pressure to *make a difference*. The institution has decided that candidates must either be outstanding on one of the following criteria, or be very strong on two or three:

- Impact on practice: evidence should cover the number of teachers and students directly affected; the nature of the improvement sought and achieved; specific expressions of interest in future development;
- Contribution to theory and/or knowledge: evidence should cover how new or synthetic the work is; warrants for trustworthiness, generality, and importance; citations; reviews; how frequently researchers elsewhere have used the ideas; and
- Improvement in either research or design methodology: evidence should cover how far the new approaches are an improvement on previous approaches; in what ways the work is robust, and applies to new situations; to what degree others employ these methods.

Given the self-interest of those who are successful under current criteria, progress in this area will not be easy; however, real leaders often have the necessary confidence to promote principled improvements. Funding agencies can play their part, as they currently do, by funding projects that require good engineering. Furthermore, they can encourage universities to give successful teams, *including their designers*, long-term appointments.

What does good engineering cost?

It is clear that the process of design and development outlined earlier (systematic design, development and evaluation) is more expensive than the simple author → publisher chain of the craft-based approach. How much does it cost? How does this investment in R&D compare with that in other fields where improvement is needed?

The NSF-funded projects for developing mathematics curriculum materials in the 1990s were each funded at a level of about $1,000,000 for each year's materials, supporting about 200 hours of teaching – that is, about $5,000 per classroom hour. Each team worked under enormous pressure to deliver at the required rate of one year's material per project year. At the Shell Centre, we have tackled less-forbidding challenges. We have developed smaller units, each supporting about 15 hours, at typical cost of £7,000–15,000 ($15,000–30,000) per classroom hour. The difference, and the range, reflects the amount of feedback, particularly classroom observation, that the funding and time has enabled. The cost of the full process (as outlined in the design, development and evaluation section, p. 132) is at the top of this range.

What would the redevelopment of the whole school curriculum cost at $30,000 per classroom hour? (No one is suggesting that everything needs to change but this gives an upper limit to the total cost.) Let us assume

- 14 years of schooling × 200 days per year × 5 hours per day = 14,000 hours
- 3 parallel developments to meet different student needs > 40,000 hours
- $30,000 per classroom hour for high-quality development

which gives a total of approximately $1.2 billion. Spread over, say, 5 years – the minimum time such a development effort would need – that yields an R&D cost of $120 million per year; since the annual expenditure on schools in the United States of America is at least $300 billion, this amounts to *investing ~0.04% of total running cost in R&D*.

Any measurable gain in the effectiveness or efficiency of schooling would justify this expenditure. (It could be saved by increasing the average number of students in a school of 2,500 students by just one student!)

For smaller countries, the proportion would be higher but still modest.

For comparison, other fields that are developing rapidly typically spend 5 per cent to 15 per cent of turnover on R&D, with 80 per cent on research-based development, 20 per cent on basic research. I believe that a level of 1 per cent of such investment in education is an appropriate target for many advanced countries. This would cover not only the R&D but the (larger) extra-implementation costs, involving as it must networks of 'live' support. This would transform the quality of children's education, with consequent benefits in personal satisfaction and economic progress.

All this takes time. However, government are used to planning and funding long-term projects in other fields – 4 years to plan and 5 years to build a bridge or an aeroplane.

Implications for policy and the design community

In this chapter, we have seen how an engineering research approach may enable

research insights ⇒ better tools and processes ⇒ improved practice

through creative design and systematic refinement using research methods. Achieving this will need changes at policy level in the strategies for educational improvement; each of these changes will depend on active effort by the research and development communities. To summarize, the strategic changes that seem to be needed are:

- Recognition that good engineering is valuable and weak engineering costly. Good engineering produces more effective and reliable outcomes, which justify the higher cost and longer time scales than the craft-based approach; persuasive evidence on this can only come from independent comparative in-depth evaluation of widely available products in use in realistic circumstances – a smarter buyer will then support better design. This needs a substantial effort by research communities, and appropriate funding.
- Coherent planning and funding of improvement by school systems, combining the long time scales of substantial educational improvement with demonstrable year-by-year gains that will satisfy political needs; the design and development community can help by linking its responses to short-term funding opportunities to a realistic long-term vision, negotiated with funders and based on basic research and past successes.
- Substantial multiskilled teams of designers, developers, evaluators and other insight researchers capable of carrying through such major projects with the long time scales they imply; while specialist centres

will continue to play an important role, there is a need for universities to play the central role they do in other 'big' fields such as physics and medicine (recognition, see above, will be important in persuading governments to make the investment needed).

- Broadening of the academic value system in universities, giving equal research credit to in-depth insights and impact on practice; this will need leadership from the research community and pressure from funders.
- Building credible theories of learning and teaching to guide research-based design and development that links to that of insight-focused research and, in turn, drives the latter to build a consensus-based core of results that are well-specified and reliable enough to be a useful basis for design.
- Collaboration: all the above will be advanced if funders, project leaders, designers and researchers learn to work more closely together over time; while the community of researchers is long established, the design and development community in education has still to acquire similar coherence.

Clearly, there is much that is challenging to be accomplished here. But, if governments and other funders become convinced that we can deliver what they need then, together, we can make educational research a more useful, more influential, and much better-funded enterprise.

Acknowledgements

The analysis in this paper owes much to discussions with Phil Daro, Alan Schoenfeld, Kaye Stacey and my colleagues at the Shell Centre – Alan Bell, Malcolm Swan, and Daniel Pead.

Notes

1 Some of the work reported in earlier chapters *is* 'engineering research' – but most of it is not. The term 'design research' covers a very wide range and I hope to show that the distinction is important.

2 The Shell Centre, also founded about 40 years ago with similar goals, has a team of about 5 people and lives from project to project on short-term funds. Sustaining long-term strategies for improvement in these circumstances requires stubborn selectivity – and luck.

3 UK university departments in all subjects undergo a Research Assessment Exercise (RAE) every 6 years. This was the definition of research for the RAE.

4 There are some areas where the need for careful *development* is recognized, notably in the development of tests; here the results have not been encouraging, largely because the *design* has often been approached on a narrow basis, dominated by traditional psychometrics, with little attention to whether what is

actually assessed truly reflects the learning goals of the subject in a balanced way. Cognitive science research and better design are both needed here.

5 The difference between Mozart and Salieri, and the hundreds of little-known composers from that time, was not in their theoretical principles; it was what each did with them. The principles and rules of melody, harmony and counterpoint were well known to, and used by, all of them.

6 The US 'What Works Clearinghouse' in its study of mathematics teaching materials, the most active area of materials development, found no study of this kind to review – this in a country with tens of thousands of educational researchers, many of them evaluators (see Schoenfeld 2006).

7 The issues of sampling, random or matched assignment need ongoing study and experiment.

8 These can also be designed to probe teachers' 'pedagogical content knowledge' of mathematics or science.

9 After many years of searching for outstanding designers, I know of only a few tens in mathematics education worldwide with whom I would be keen to work.

10 It is true that some educational software is developed somewhat more systematically – the inevitably high cost of design and programming makes room for this. However, even this is held back by the same lack of reliable data for users on how well it works.

11 The 'What Works Clearinghouse' in the US has such a purpose; the methodology is profoundly flawed (see Schoenfeld (2006)) – but perhaps it is a start.

References

Bell, A. (1993). Some experiments in diagnostic teaching. *Educational Studies in Mathematics*, *24*(1), 115–37. Also URL: www.toolkitforchange.org

Burkhardt, H. and Schoenfeld, A. H. (2003). Improving educational research: Towards a more useful, more influential and better funded enterprise. *Educational Researcher*, *32*(9), 3–14.

Cockcroft Report (1979). *Mathematics Counts*. London: HMSO.

Gardner, H. and Shulman, L. S. (2005). *The Professions in America Today* (and other articles on *Professions and Professionals*). Daedalus, Summer 2005.

HEFC (1999). *Guidance on Submissions to the Research Assessment Exercise* (paragraph 1.12). London: Higher Education Funding Council for England and Wales 1999. URL: www.hero.ac.uk/rae/

OECD (2003). *The PISA 2003 assessment framework: Mathematics, reading science and problem solving knowledge and skills*, Paris. URL: https://www.pisa.oecd.org/dataoecd/38/51/33707192.pdf

RAE (Research Assessment Exercise) (2001). Research assessment exercise 2001: Briefing notes. URL: http://195.194.167.103/Pubs/2_01

Schoenfeld, A. H. (2002). Research methods in (mathematics) education. In L. English (ed.), *Handbook of International Research in Mathematics Education* (pp. 435–88). Mahwah, NJ: Erlbaum.

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the what works clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher 35*(2), 13–21.

Shell Centre (1987–89), Swan, M., Binns, B., Gillespie, J., and Burkhardt, H., *Numeracy Through Problem Solving* (five modules for curriculum and assessment in mathematical literacy). Harlow: Longman, revised Shell Centre Publications, Nottingham, UK, 2000. URL: www.mathshell.com/scp/index.htm

Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technical Innovation*. Washington, DC: Brookings.

Tomlinson Report (2004). *14–19 Curriculum and Qualifications Reform*. Department for Education and Skills. London: HMSO. URL: www.14–19reform.gov.uk