EDUCATIONAL DESIGNER
JOURNAL OF THE INTERNATIONAL SOCIETY FOR DESIGN AND DEVELOPMENT IN EDUCATION

On Strategic Design

Hugh Burkhardt

Abstract

This paper develops the concept of "strategic design", the design implications of the interactions of a product with the whole user system, and relates it to other aspects of design. It describes some examples of poor strategic design that occur frequently, and some cases where effective strategic design has been important in the large-scale impact of an ambitious educational innovation. From these, the paper then seeks to infer some principles for strategic design. It is aimed at the three major constituencies of ISDDE: designers, design team leaders, and the client-funders that often commission their work. The hope is that sharpened awareness of the importance, and the challenges, of strategic design may help to increase the impact of good design as a whole.

The goals of this paper are to:

- Articulate and illustrate the concept of "strategic design";
- Relate it to other aspects of design;
- Pursue some of the issues it raises; and
- Suggest some principles for strategic design.

I will start with a working description:

Strategic design focuses on the design implications of the interactions of the products, and the processes for their use, with the whole user system it aims to serve.

The importance of strategic design is illustrated by the many examples of design excellence that have been undermined by poor strategic design – wonderful lessons, assessment tasks, and professional development activities that are never seen[1], while mediocrity (and worse) is widespread.

Section 1 outlines the concept, which will be developed throughout the paper, distinguishing *strategic, tactical* and *technical* design. Section 2 illustrates the concept with three areas of poor strategic design that are commonplace across education systems, and the design challenges each presents. Section 3 sets out to identify underlying causes of poor strategic design, and the contributions to it of client-funders, education professionals, and poor methodology. Switching to a more upbeat note, Section 4 describes four projects that paid close attention to strategic design and that have had substantial impact on the systems they aimed to improve. The final sections set out some principles for strategic design, issues that need further study, long and medium term goals, and immediate actions that can forward their achievement.

1. Aspects of educational design

I choose here to distinguish three major aspects of educational design – *strategic, tactical* and *technical*[2] – if only to make clear what strategic design is *not*. (Illustrative examples are given in parentheses below.)

Technical design is the detailed process with which *any* designer is familiar. It is focused on the design of individual elements of the product (e.g. a teaching unit; a professional development module; an assessment task). Technical design is focused on the end users and their environment (students and the teacher in classrooms; teachers in professional development activities; the diverse students taking a test, and those who will score their responses).

Technical design is the responsibility of the *lead designer* of the unit.

Tactical design is focused on the overall internal structure of the product (e.g. a multi-year set of teaching materials; a year's assessment; a professional development package). Typically it involves such things as:

- Specification of core design principles, selected in the light of prior research on learning, teaching, and/or professional development trajectories – or, too often, just marketing;
- Selection of specific learning and performance goals, including strands of progression;
- Specifying sequences and cross-connections within the materials, balancing linear coherence with diverse multiple connections (among concepts and contexts, standard results to learn and open investigations to experience).

Tactical design is a responsibility of the design team leaders and lead designers, working with their colleagues in the design team.

Strategic design, the focus of this paper, is concerned with the overall structure of the product set and how it will relate to the user-system. It applies in different forms to most of the products and processes that educational designers tackle: curriculum specifications; assessment; teaching materials; professional development processes and materials; building system capacity in various ways. Typically strategic design involves not simply the end-users (e.g. teachers and their students) but all the key communities involved who will affect decisions on the framework within which the users work – school leadership; school system leadership; politicians; parents; and various other professions, such as assessment designers and researchers.

Strategic design includes such things as:

- Identifying a specific opportunity for improvement;
- Selecting a set of improvement goals;
- Designing the overall structure of a set of tools that can forward them;
- Choosing or designing a model of change (whether, for example, comprehensive or more specific; one-step or gradual; curriculum-led, assessment-led, or professional development led) along with the phases, pacing and timing of implementation;
- Identifying the resources that are needed to do the job well (how much design effort, trialling, implementation support, and of what kinds), and the compromises that are acceptable;
- Recognizing and questioning constraints from the client's *grand strategy* (generic performance goals; alignment; model of change; top-down v proposal driven); and
- Advising the client on the likely implications of their various decisions, including their likely unintended consequences and uncertainties – and suggesting changes.

Strategic design is a responsibility of the design team leadership, usually in negotiation with the client-funder – often government, a quasi-government agency, or a foundation.

There is no hierarchy of importance among strategic, tactical and technical design. While this paper focuses on strategic design, all three are important if the product is to work well. All three offer opportunities for *creativity in the search for excellence* – and for making ghastly choices that undermine the whole enterprise.

My own view is that poor *strategic design* is the most common cause of *failure*, while excellence in *technical design* is the main source of the magic combination of *power, surprise and delight* that characterizes really outstanding products[3] – as in music, art and literature, *details matter*. *Tactical design* is central to the coherence of the enterprise.

This framework complements Goodlad's (1994) rather different analytic perspective on curriculum design, which distinguishes:

1. *The socio-political perspective* – the influence exercised by various individual and organizational stakeholders;
2. *The technical-professional perspective* – the methods of the curriculum development process;
3. *The substantive perspective* – the question of what should be learned[4].

This paper suggests that 1 must be part of 2, explicitly seen as part of the design and development process and proactively addressed as such.

Design control is the other concept that belongs here. How are design decisions made, and by whom? While all members of a design team will contribute ideas and suggestions on all three aspects of design, it is worth identifying how *choices* are made among the huge range of possibilities that any design task affords. The obvious principle is to make the best use of the diverse design talent available in the team. This hierarchy of decision taking will influence the design and its impact.

There are various approaches to design control. Some small teams work by consensus – this has obvious advantages but can lead to long unproductive discussions and suboptimal compromises [5]. In contrast, as in architecture, design control may rest with a single lead designer – or, sometimes, a small group who have worked closely together for a long time. Alternatively, different people may have design control over different aspects of the design, reflecting their strengths – e.g. as strategic, tactical or technical designers, as software designers, or in relation to specific learning goals.

Whatever the choice, I have found that it is important to make clear the locus of design control – this smooths and speeds the design process, while leaving most room for individual design flair.

2. Common failures in strategic design

More often than not, educational initiatives that seek to improve student learning fail to achieve their stated goals. This paper makes the case that this is often, at least partly, due to poor strategic design. This is unsurprising. Strategic design is often assigned to committees of advisers by the client-funder, with both seeing it as a policy issue, rather than a design challenge that may be crucial to the success of the innovation. For government agencies, ad hoc decisions, dominated by practical policy considerations, are the norm. If we are to do better, we must understand the phenomena involved. I begin with some examples, moving on in the next section to look at underlying causes.

Those who seek examples of poor strategic design face *un embaras de richesse*. Many initiatives have doom written all over them – predictable, and often predicted. Some ignore well-known features of the system – for example, that most teachers teach to the test when high stakes are involved. Some fail to recognize that a design does not reflect its purpose – for example, that specifying performance goals involves more than a list of topics in mathematics or science. Some show no sign of any systematic attempt to reconcile their usually-ambitious goals with the limitations of the process chosen for achieving them – for example, that a few sessions of discussion will not enable teachers to profoundly expand their range of classroom teaching skills. The following three examples are all repeated regularly in many countries and school systems. The outline of each that follows focuses on its strategic design, the form of its failure, and the design challenges that must be overcome if we are to do better.

2A Assessment – the "only measurement" fallacy

Policy makers in the Anglophone countries and some others are wedded to using tests of various kinds as prime instruments of system control. Tests are seen as reliable measures[6] of student, teacher and school performance, forming the basis of each school's "accountability" to the society that funds it. Targets are set in terms of test scores that have serious consequences for those concerned. Students' access to higher education depends on their test scores. In England, schools are ranked on test scores into "league tables" to guide "parental choice[7]". Schools that under-perform may be "taken into special measures" or closed. Similar sanctions apply in the US.

Given the importance of tests, it seems obvious that their design should be a focus of attention. They should embody the full set of performance goals in a balanced way[8]. Yet this central responsibility of test providers and those that commission test design is widely ignored, and sometimes denied. Their focus is on the statistical properties of the test and the "fairness" of the procedures, with little attention to what aspects of performance are assessed[9]. Policy makers talk and behave as though tests are just "measurement"; they choose simple tests because they are cheap and, if pressed, argue that the results correlate with more valid and elaborate assessments. Most articulate education professionals dislike tests so much that, hoping to marginalize testing, they make no serious effort to improve the current versions.

This approach ignores two of the three roles that high-stakes assessment *inevitably* plays. In brief, it:

A. *Measures levels of student performance,* but only across the range of task-types used;

B. *Exemplifies performance objectives* – the types of task in high-stakes tests show what kinds of performance will be recognized and rewarded in a clear form that teachers and students readily understand; as a result, this set of task types

C. *Dominates classroom activities* – the task types in high-stakes tests largely determine the pattern of teaching and learning activities in most classrooms.

Thus assessment design is the unnoticed "elephant in the room" in the planning of improvement programs. There is plenty of evidence that "what you test is what you get" *(WYTIWYG)* is a fact of life (see e.g. Black and Atkin 1996, Barnes, Clarke and Stephen 2000). So in systems with high-stakes assessment, the tests *are* the *de facto* standards. While the UK national inspectors of schools (Ofsted 2006, 2008) remark with regret on the dominance of test-focused activities, teachers regard it as inevitable – after all, these are the measures of their performance that society has decided to value. More hopefully, where balanced high-stakes tests have been adopted, they have proven to be a powerful influence in improving teaching and learning in classrooms (see Section 4).

The design challenge

The design of well-balanced assessment in a form that can be used for accountability purposes has been a solved problem for many years. There are working examples around the world of high-stakes timed examinations that show what can be done, and how it can enhance learning. They are not perfect but are vastly better balanced than most current tests. History contains many outstanding examinations that *enabled students to show what they know, understand and can do*[10]. The strategic design principle here is to include task types that represent the full range of performance goals.

The cost and complexity of high-quality balanced assessment is greater than for machine-marked multiple choice tests; more complex tasks cannot be set and scored for $1 per student-test, a widely-accepted cost target in the US. (The massive cost of the class time wasted on otherwise-unproductive test-prep is generally ignored.)

There are also well-established ways of lowering the cost of assessment so that it can monitor standards as reliably as at present, while *enhancing* student learning. A strategy that has multiple benefits is to make teachers the prime assessors, providing them with good assessment tools and some training, and monitoring their scoring on a sampling basis. The many examples of this approach in practice show that it is also powerful professional development for the teachers involved. It links naturally to formative assessment in the classroom, which research shows to be such a powerful way of improving learning (Black and Wiliam 1998).

Strategically, it is actually unwise to hold costs for structured assessment down to current levels, well below 1% of the ~$10,000 per student-year that education typically costs. Feedback is crucial factor in determining the behaviour of systems of all kinds. Well–structured feedback on student achievement (Role A above), performance goals (Role B), and exemplar tasks for the classroom (Role C) are worth far more than the current investments in these areas.

Even when research-based methods of design and development have been used in assessment, notably in some test development, the commissioning specification has often been too narrow, excluding design solutions that would allow the realization of the policy goals. The purely statistical methods used in traditional psychometrics inevitably move attention from the kinds of performances that are assessed, which vary from subject to subject, to the statistical properties of the test[11].

2B How "standards" drive down standards

Many current models of national and state curriculum specifications ("standards" in what follows) in mathematics and science are examples of bad strategic design – they have the effect in practice opposite to that intended. They actually drive down standards of performance in the subject. In explaining this I shall use as the lead example the National Curriculum for Mathematics in England. However, many state standards in the US and elsewhere have much the same structure – and effect.

Criterion referencing is the source of the problem. The National Curriculum and most current mathematics "standards" in the US were designed on the principle that achievement goals can be specified through a detailed list of *level criteria* – concepts and skills that a student at that level should know, understand and show in tests. For example:

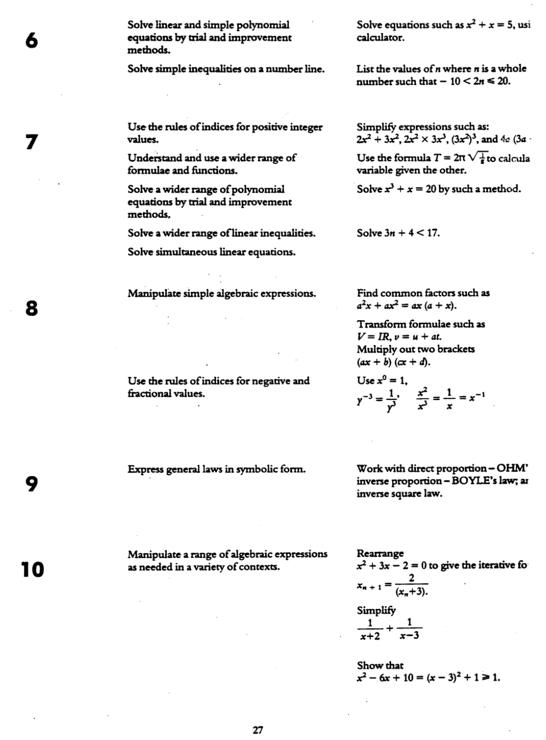
> *Use the rules of indices for positive integer values, e.g.*
> *simplify expressions such as* $2x^2 + 3x^2$, $2x^2 \times 3x^2$, $(3x^2)^3$
>
> From Level 7 of 1988 UK National Curriculum design: Algebra Target 2 (see Figure 1)

Or:

> *Factor simple quadratic expressions with integer coefficients, e.g.*
> $x^2 + 6x + 9$, $x^2 + 2x - 3$, *and* $x^2 - 4$;
>
> *Solve simple quadratic equations, e.g.*
> $x^2 = 16$ *or* $x^2 = 5$ *(by taking square roots);* $x^2 - x - 6 = 0$, $x^2 - 2x = 15$ *(by factoring);*
>
> *verify solutions by evaluation.*
>
> From Michigan Grade Level Content Expectations - Grade 8 Algebra item A.FO.08.08 (See Figure 2).

Note the brevity of the task examples given.

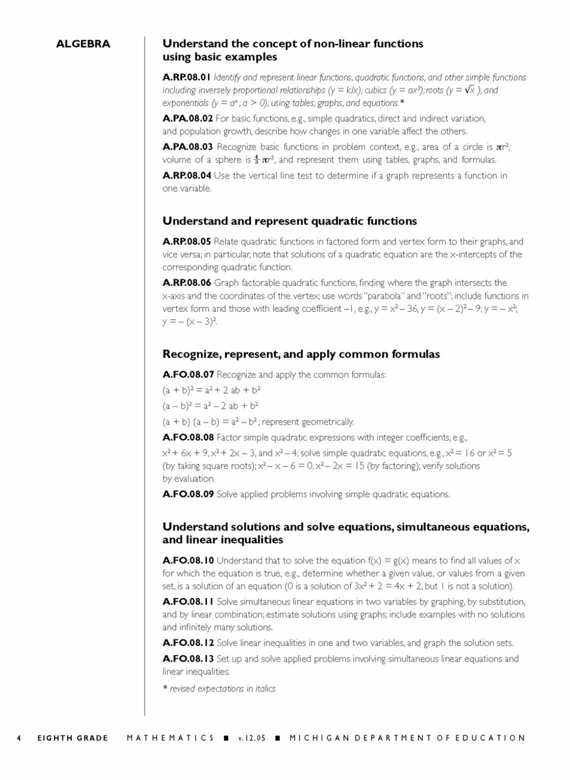
Figure 1: Example of *level criteria* from the 1988 UK National Curriculum design



6	Solve linear and simple polynomial equations by trial and improvement methods.	Solve equations such as $x^2 + x = 5$, usi calculator.
	Solve simple inequalities on a number line.	List the values of n where n is a whole number such that $-10 < 2n \leqslant 20$.
7	Use the rules of indices for positive integer values.	Simplify expressions such as: $2x^2 + 3x^2$, $2x^2 \times 3x^3$, $(3x^2)^3$, and 4a (3a
	Understand and use a wider range of formulae and functions.	Use the formula $T = 2\pi\sqrt{\frac{l}{g}}$ to calcula variable given the other.
	Solve a wider range of polynomial equations by trial and improvement methods.	Solve $x^3 + x = 20$ by such a method.
	Solve a wider range of linear inequalities.	Solve $3n + 4 < 17$.
	Solve simultaneous linear equations.	
8	Manipulate simple algebraic expressions.	Find common factors such as $a^2x + ax^2 = ax(a + x)$. Transform formulae such as $V = IR$, $v = u + at$. Multiply out two brackets $(ax + b)(cx + d)$.
	Use the rules of indices for negative and fractional values.	Use $x^0 = 1$, $y^{-3} = \frac{1}{y^3}$, $\frac{x^2}{x^3} = \frac{1}{x} = x^{-1}$
9	Express general laws in symbolic form.	Work with direct proportion – OHM' inverse proportion – BOYLE's law; ar inverse square law.
10	Manipulate a range of algebraic expressions as needed in a variety of contexts.	Rearrange $x^2 + 3x - 2 = 0$ to give the iterative fo $x_{n+1} = \frac{2}{(x_n+3)}$. Simplify $\frac{1}{x+2} + \frac{1}{x-3}$ Show that $x^2 - 6x + 10 = (x - 3)^2 + 1 \geqslant 1$.

27

Figure 2: Michigan *Grade Level Content Expectations* - Grade 8 Algebra



ALGEBRA

Understand the concept of non-linear functions using basic examples

A.RP.08.01 *Identify and represent linear functions, quadratic functions, and other simple functions including inversely proportional relationships ($y = k/x$); cubics ($y = ax^3$); roots ($y = \sqrt{x}$); and exponentials ($y = a^x$, $a > 0$); using tables, graphs, and equations.* *

A.PA.08.02 For basic functions, e.g., simple quadratics, direct and indirect variation, and population growth, describe how changes in one variable affect the others.

A.PA.08.03 Recognize basic functions in problem context, e.g., area of a circle is πr^2, volume of a sphere is $\frac{4}{3}\pi r^3$, and represent them using tables, graphs, and formulas.

A.RP.08.04 Use the vertical line test to determine if a graph represents a function in one variable.

Understand and represent quadratic functions

A.RP.08.05 Relate quadratic functions in factored form and vertex form to their graphs, and vice versa; in particular, note that solutions of a quadratic equation are the x-intercepts of the corresponding quadratic function.

A.RP.08.06 Graph factorable quadratic functions, finding where the graph intersects the x-axis and the coordinates of the vertex; use words "parabola" and "roots"; include functions in vertex form and those with leading coefficient –1, e.g., $y = x^2 - 36$, $y = (x - 2)^2 - 9$; $y = -x^2$; $y = -(x - 3)^2$.

Recognize, represent, and apply common formulas

A.FO.08.07 Recognize and apply the common formulas:
$(a + b)^2 = a^2 + 2ab + b^2$
$(a - b)^2 = a^2 - 2ab + b^2$
$(a + b)(a - b) = a^2 - b^2$; represent geometrically.

A.FO.08.08 Factor simple quadratic expressions with integer coefficients, e.g., $x^2 + 6x + 9$, $x^2 + 2x - 3$, and $x^2 - 4$; solve simple quadratic equations, e.g., $x^2 = 16$ or $x^2 = 5$ (by taking square roots); $x^2 - x - 6 = 0$, $x^2 - 2x = 15$ (by factoring); verify solutions by evaluation.

A.FO.08.09 Solve applied problems involving simple quadratic equations.

Understand solutions and solve equations, simultaneous equations, and linear inequalities

A.FO.08.10 Understand that to solve the equation $f(x) = g(x)$ means to find all values of x for which the equation is true, e.g., determine whether a given value, or values from a given set, is a solution of an equation (0 is a solution of $3x^2 + 2 = 4x + 2$, but 1 is not a solution).

A.FO.08.11 Solve simultaneous linear equations in two variables by graphing, by substitution, and by linear combination; estimate solutions using graphs; include examples with no solutions and infinitely many solutions.

A.FO.08.12 Solve linear inequalities in one and two variables, and graph the solution sets.

A.FO.08.13 Set up and solve applied problems involving simultaneous linear equations and linear inequalities.

* *revised expectations in italics*

4 EIGHTH GRADE MATHEMATICS ■ v.12.05 ■ MICHIGAN DEPARTMENT OF EDUCATION

Criterion referencing is an attractively simple idea. The public accepts it and policy makers on both sides of the Atlantic seem to love it[12]. But it is a dangerous illusion. What is the problem? Fundamentally, it is that:

> *The level of difficulty of a substantial task depends on various interacting factors – increasing with the complexity, unfamiliarity, and technical demand of the task, and the autonomy expected of the student in tackling it.*

Thus the difficulty of the task is higher than that of its technical elements, tested separately – a rich task that is challenging for a good 16 year-old student (called level 7) may require only mathematical concepts and skills that were taught in elementary school (level 4 and below). The "Consecutive Sums" task is an example[13].

Consecutive sums

The number 9 can be written as the sum of consecutive whole numbers in two ways:

9 = 2 + 3 + 4

9 = 4 + 5

The number 16 cannot be written as a consecutive sum.

Now look at other numbers and find out all you can about writing them as sums of consecutive whole numbers.

OK, this seems fairly obvious – but why are criterion-based standards *dangerous*? Because, it is only *fair* to give students the opportunity to meet the criteria for the highest level they might be able to reach – this is achieved by testing each concept and skill separately with a short topic-focused item that has no other *cognitive load* (from complexity, unfamiliarity, or longer chains of reasoning) that would increase its difficulty. In the following task (from Grade 10 GCSE):

(a) Factorise $x^2 - 10x + 21$
(b) Hence solve $x^2 - 10x + 21 = 0$

..note the fragmentation of an already straightforward exercise; this is done to test explicitly the two criteria:

- Can factorise a quadratic expression
- Can solve a quadratic equation

This approach is the only way that “standards” which define levels through detailed lists of concepts and skills can be made to work. UK mathematics tests now consist of that kind of fragmented performance which, because the stakes are high, also dominates classroom learning activities (Ofsted 2006, 2008) . Clear evidence that such fragmentation is commonplace can be found by comparing test items with standards, as above.

The damage to student learning is profound. Success with such fragments has little value outside the mathematics classroom; it surely does not guarantee success with the more substantial chains of reasoning that doing and using mathematics involves. To be useful in solving substantial problems, from the real world or within mathematics, a technique needs multiple connections in the student's mind – to other math concepts and to diverse problem contexts within and outside mathematics. These connections are built over time, by learning how to tackle more complex tasks like *Consecutive Sums*. Such tasks (see Figure 3 for more examples) are much more challenging than their technical demand suggests because the strategic demand is a major part of the *total cognitive load* that determines difficulty.

To summarise, when "standards" are based on criterion referencing by topic, the level criteria inevitably (on grounds of fairness) require short item testing focused on the listed topics, which leads to short item teaching (via WYTIWYG, explained in section 2A).

This range of task-types covers only a narrow subset of performance goals that is useless outside schools. This undermines student learning by not preparing students to *think with mathematics* about the more substantial tasks they will meet in life outside the classroom – the epitome of low standards.

Figure 3: Assessment task exemplars (sample)

Animals

This problem gives you the chance to:
- find the median, mode and range of a set of tabulated data
- interpret graphs

Hugh works in a nature reserve. He has caught a number of animals of one species in an area of woodland. He records their weights in grams, then he puts their weights in grouped frequency tables.

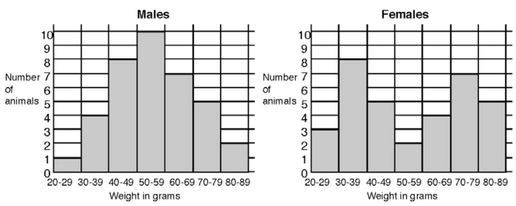
Males

Weight in grams	Number of animals
20 - 29	1
30 -39	4
40 - 49	8
50 - 59	10
60 - 69	7
70 - 79	5
80 - 89	2

Females

Weight in grams	Number of animals
20 - 29	3
30 -39	8
40 - 49	5
50 - 59	2
60 - 69	4
70 - 79	7
80 - 89	5

Hugh uses these tables to plot the two graphs shown below.



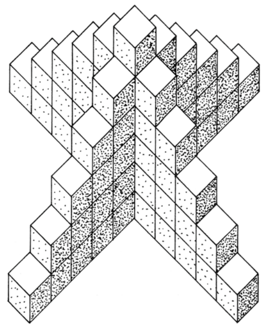
1. Find the median weights of the male and female animals. In which groups do they lie?

2. Using the graphs and tables for evidence, write three different statements comparing the similarities and differences between the weights of the male and female animals.

Figure 3 (sample, continued)

Download as PDF

SKELETON TOWER



1. How many cubes are needed to build this tower?
2. How many cubes are needed to build a tower like this, but 12 cubes high? Explain how you work out your answer.
3. How would you calculate the number of cubes needed for a tower n cubes high?

(This is an extract from a larger collection available with the online version of this article)

The design challenge

How might one design "standards" that set clear learning and performance goals without narrowing the curriculum? There have been various attempts at improving criteria to include strategic and tactical skills (often called *processes*) at different levels. There is a fundamental problem here too: the same strategies and tactics help people solve problems across the range of difficulty. Again, *it is tasks, not processes, that have well-defined "levels" of difficulty.*

Other countries have taken a quite different approach to the design of standards in mathematics and science, describing the learning and performance goals in broad terms. This approach relies on the professional expertise of teachers and others to find a more detailed realisation that is appropriate to their local circumstances. The "flower diagram" (Figure 4) used in mathematics standards in Denmark illustrates this approach. These broad descriptions of competencies do not define levels of difficulty. So it is not surprising that they are common in school systems that do not use tests as an accountability tool with high-stakes consequences. However, it was also common in traditional British examinations, where the experienced task designer recognized the various aspects of challenge in a task and adjusted the overall difficulty appropriately[14].

One key to better design is to recognize the importance of tasks in defining standards. Specific task exemplars, complemented by examples of various levels of student work on the task, communicate learning and performance goals in a form that everyone understands[15].

Figure 4: The Danish "flower diagram"

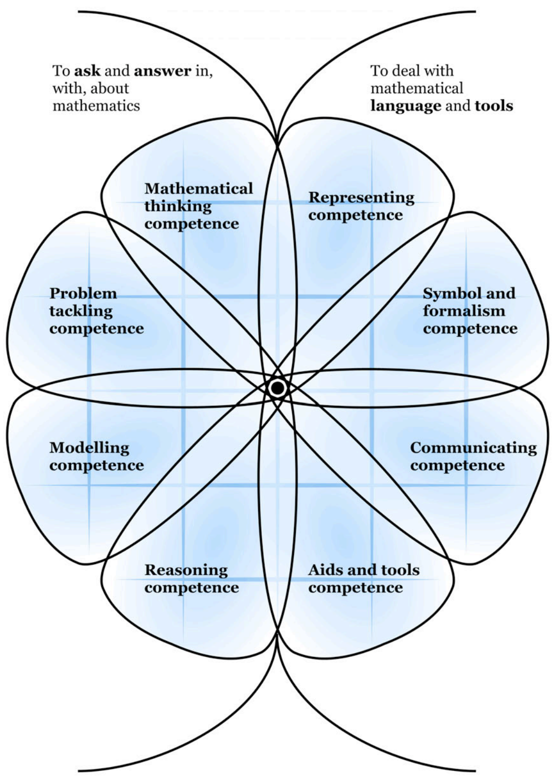


Figure 4 (continued)

Asking and answering questions	Mathematical language and tools
Mathematical thinking competence such as • *posing questions* that are characteristic of mathematics, and *knowing the kinds* of answers (not necessarily the answers themselves or how to obtain them) that mathematics may offer; • understanding and handling the *scope* and *limitations* of a given *concept*. • *extending* the scope of a *concept* by *abstracting* some of its properties; *generalizing results* to larger classes of objects; • *distinguishing* between different *kinds of mathematical statements* ('if-then', assumptions, definitions, theorems, conjectures, cases)	**Representing competence** such as • *understanding* and *utilizing* (decoding, interpreting, distinguishing between) different sorts of representations of mathematical objects, phenomena and situations; • understanding and utilizing the *relations between different representations* of the same entity, including knowing about their relative strengths and limitations; • *choosing* and *switching* between representations.
Problem solving competence such as • *identifying*, *posing*, and *specifying* different kinds of mathematical *problems* – pure or applied; open-ended or closed; • *solving* different kinds of mathematical problems (pure or applied, open-ended or closed), whether posed by others or by oneself, and, if appropriate, in different ways.	**Symbol and formalism competence** such as • *decoding* and *interpreting symbolic and formal* mathematical *language*, and understanding *its relations to natural language*; • understanding the *nature* and *rules* of *formal mathematical systems* (both syntax and semantics); • *translating* from *natural language* to *formal/symbolic language* *handling* and manipulating statements and *expressions* containing *symbols* and *formulae*.
Modeling competence such as • *analyzing* foundations and properties of *existing models*, including assessing their range and validity • *decoding* existing models, i.e. translating and interpreting model elements in terms of the 'reality' modeled • *active modeling* in a given context: structuring the field; mathematising; working with(in) the model, including solving the problems it gives rise to; validating the model, internally and externally; analyzing and criticizing the model, in itself and vis-à-vis possible alternatives; communicating about the model and its results; monitoring and controlling the entire modeling process.	**Communicating competence** such as • *understanding others'* written, visual or oral 'texts', in a variety of linguistic registers, about matters having a mathematical content; • *expressing oneself*, at different levels of theoretical and technical precision, in oral, visual or written form, about such matters.
Reasoning competence such as • *following* and *assessing chains of arguments*, put forward by others • *knowing* what a mathematical *proof is* (not), ands how it differs from other kinds of mathematical reasoning, e.g. heuristics • *uncovering* the *basic ideas* in a given line of argument (especially a proof), including distinguishing main lines from details, ideas from technicalities; *devising* formal and informal mathematical *arguments*, and *transforming* heuristic arguments to valid proofs, i.e. *proving statements.*	**Aids and tools competence** such as • *knowing* the *existence* and *properties* of various tools and aids for mathematical activity, and their range and limitations; • being able to *reflectively use* such aids and tools, including IT.

Since *difficulty is a property of the task, not its separate elements,* it can only be reliably determined by trialling the task with students, and recognizing student responses at different levels in the scoring scheme. Thus any valid level scheme should be based on a set of well-analyzed tasks to which other tasks can then be related through trialling.

In an earlier paper, *On specifying a curriculum* (Burkhardt 1990), prepared in the light of experience during the design of the British National Curriculum, I pointed out that the final version gave no indication as to the types and balance of tasks that were to represent the performance goals in Mathematics[16] – the concepts and skills could be shown entirely in short items, or in the course of three week-long projects, or in a variety of other task types in between. I argued that to specify a curriculum relatively unambiguously, you need three *independent* elements (see Figure 5):

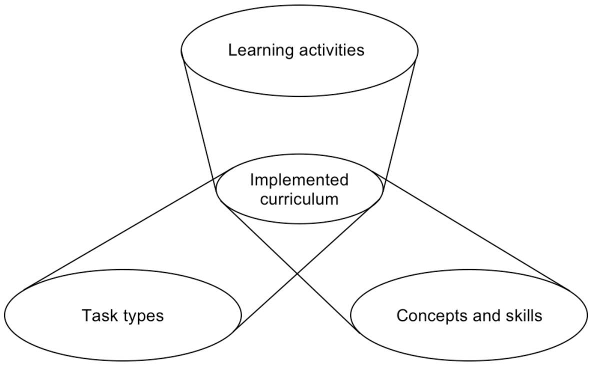


Figure 5: Three dimensions in specifying a curriculum

- The tools in the toolkit of mathematical *concepts and skills*
- The performance targets, as exemplified by *task types*
- The pattern of classroom *learning activities*

They are independent, in that none of them determines the others, and complementary, each supporting the others.

Currently in both the UK and the US there are attempts to produce improved models of standards. The extract shown in Figure 6 is from the 2008 standards of the Qualifications and Curriculum Development Agency (QCDA) in England (QCDA 2007). Note the general descriptions of processes and the partial move away from detailed lists of techniques; but it is clear that any of the criteria can be interpreted at very different levels of difficulty. The tendency to narrow the task set remains – the easiest way to test, say, the process of *representation* is separately, not as part of solving a substantial non-routine problem. Since the processes do not change much across ages and levels – it is easy to find tasks that a typical 7 year old can do (~Level 2) that involve these processes –the focus tends to remain on the content descriptions at each level (Figure 6, page 2).

Currently in the US, a different kind of model is being developed for the draft “College and Career Readiness Standards for Mathematics”, commissioned by the Governors of US states as model national standards (NGA, CCSSO 2009). This draft describes mathematical practices and principles in broad terms (see Figure 7). Notably, it avoids detailed lists of technique, replacing them with a relatively rich set of tasks (see Figure 7, page 2), covering a broad variety of task types, that exemplifies the range of performance being sought. Its progress through the dynamics of each state’s education policy formation will be interesting.

Figure 6: Extracts from English National Curriculum

Attainment targets

Attainment target 1: Mathematical processes and applications

Level 4
Pupils develop their own strategies for solving problems and use these strategies both in working within mathematics and in applying mathematics to practical contexts. When solving problems, with or without a calculator, they check their results are reasonable by considering the context or the size of the numbers. They look for patterns and relationships, presenting information and results in a clear and organised way. They search for a solution by trying out ideas of their own.

Level 5
In order to explore mathematical situations, carry out tasks or tackle problems, pupils identify the mathematical aspects and obtain necessary information. They calculate accurately, using ICT where appropriate. They check their working and results, considering whether these are sensible. They show understanding of situations by describing them mathematically using symbols, words and diagrams. They draw simple conclusions of their own and explain their reasoning.

Level 6
Pupils carry out substantial tasks and solve quite complex problems by independently and systematically breaking them down into smaller, more manageable tasks. They interpret, discuss and synthesise information presented in a variety of mathematical forms, relating findings to the original context. Their written and spoken language explains and informs their use of diagrams. They begin to give mathematical justifications, making connections between the current situation and situations they have encountered before.

Level 7
Starting from problems or contexts that have been presented to them, pupils explore the effects of varying values and look for invariance in models and representations, working with and without ICT. They progressively refine or extend the mathematics used, giving reasons for their choice of mathematical presentation and explaining features they have selected. They justify their generalisations, arguments or solutions, looking for equivalence to different problems with similar structures. They appreciate the difference between mathematical explanation and experimental evidence.

Level 8
Pupils develop and follow alternative approaches. They compare and evaluate representations of a situation, introducing and using a range of mathematical techniques. They reflect on their own lines of enquiry when exploring mathematical tasks. They communicate mathematical or statistical meaning to different audiences through precise and consistent use of symbols that is sustained throughout the work. They examine generalisations or solutions reached in an activity and make further progress in the activity as a result. They comment constructively on the reasoning and logic, the process employed and the results obtained.

Exceptional performance
Pupils critically examine the strategies adopted when investigating within mathematics itself or when using mathematics to analyse tasks. They explain why different strategies were used, considering the elegance and efficiency of alternative lines of enquiry or procedures. They apply the mathematics they know in a wide range of familiar and unfamiliar contexts. They use mathematical language and symbols effectively in presenting a convincing, reasoned argument. Their reports include mathematical justifications, distinguishing between evidence and proof and explaining their solutions to problems involving a number of features or variables.

Attainment target 2: Number and algebra

Level 4
Pupils use their understanding of place value to multiply and divide whole numbers by 10 or 100. When solving number problems, they use a range of mental methods of computation with the four operations, including mental recall of multiplication facts up to 10 × 10 and quick derivation of corresponding division facts. They use efficient written methods of addition and subtraction and of short multiplication and division. They recognise approximate proportions of a whole and use simple fractions and percentages to describe these. They begin to use simple formulae expressed in words.

Level 5
Pupils use their understanding of place value to multiply and divide whole numbers and decimals. They order, add and subtract negative numbers in context. They use all four operations with decimals to two places. They solve simple problems involving ratio and direct proportion. They calculate fractional or percentage parts of quantities and measurements, using a calculator where appropriate. They construct, express in symbolic form and use simple formulae involving one or two operations. They use brackets appropriately. They use and interpret coordinates in all four quadrants.

Attainment target 3: Geometry and measures

Level 4
Pupils make 3D mathematical models by linking given faces or edges, and draw common 2D shapes in different orientations on grids. They reflect simple shapes in a mirror line. They choose and use appropriate units and tools, interpreting, with appropriate accuracy, numbers on a range of measuring instruments. They find perimeters of simple shapes and find areas by counting squares.

Level 5
When constructing models and drawing or using shapes, pupils measure and draw angles to the nearest degree and use language associated with angles. They know the angle sum of a triangle and that of angles at a point. They identify all the symmetries of 2D shapes. They convert one metric unit to another. They make sensible estimates of a range of measures in relation to everyday situations. They understand and use the formula for the area of a rectangle.

Attainment target 4: Handling data

Level 4
Pupils collect discrete data and record them using a frequency table. They understand and use the mode and range to describe sets of data. They group data in equal class intervals where appropriate, represent collected data in frequency diagrams and interpret such diagrams. They construct and interpret simple line graphs.

Level 5
Pupils understand and use the mean of discrete data. They compare two simple distributions using the range and one of the mode, median or mean. They interpret graphs and diagrams, including pie charts, and draw conclusions. They understand and use the probability scale from 0 to 1. They find and justify probabilities and approximations to these by selecting and using methods based on equally likely outcomes and experimental evidence, as appropriate. They understand that different outcomes may result from repeating an experiment.

Figure 7: Extracts from the US *College and Career Readiness Standards*

7. **Modeling** | see evidence
Modeling uses mathematics to help us make sense of the real world—to understand quantitative relationships, make predictions, and propose solutions.

A model can be very simple, such as a geometric shape to describe a physical object like a coin. Even so simple a model involves making choices. It is up to us whether to model the solid nature of the coin with a three-dimensional cylinder, or whether a two-dimensional disk works well enough for our purposes. For some purposes, we might even choose to adjust the right circular cylinder to model more closely the way the coin deviates from the cylinder.

In any given situation, the model we devise depends on a number of factors: How precise an answer do we want or need? What aspects of the situation do we most need to understand, control, or optimize? What resources of time and tools do we have? The range of models we can create and analyze is constrained as well by the limitations of our mathematical and technical skills. For example, modeling a physical object, a delivery route, a production schedule, or a comparison of loan amortizations each requires different sets of tools. Networks, spreadsheets and algebra are powerful tools for understanding and solving problems drawn from different types of real-world situations. One of the insights provided by mathematical modeling is that essentially the same mathematical structure might model seemingly different situations.

The basic modeling cycle is one of (1) identifying the key features of a situation, (2) creating geometric, algebraic or statistical objects that describe key features of the situation, (3) analyzing and performing operations on these objects to draw conclusions and (4) interpreting the results of the mathematics in terms of the original situation. Choices and assumptions are present throughout this cycle.

Connections to Quantity, Equations, Functions, Shape, Coordinates and Statistics. Modeling makes use of shape, data, graphs, equations and functions to represent real-world quantities and situations.

Core Concepts
Students understand that:

A. Mathematical models involve choices and assumptions that abstract key features from situations to help us solve problems.
see examples

B. Even very simple models can be useful.
see examples

Core Skills
Students can and do:

1. Model numerical situations.
see examples
2. Model physical objects with geometric shapes.
see examples
3. Model situations with equations and inequalities.
see examples
4. Model situations with common functions.
see examples
5. Model situations using probability and statistics.
see examples
6. Interpret the results of applying a model and compare models for a particular situation.
see examples

Example Tasks

1. **Core Concept A; Core Skill 2.**
If everyone in the world went swimming in Lake Michigan, what would happen to the water level? (Would Chicago be flooded?)
2. **Core Concept A; Core Skill 2; Core Skill 7.**
The Federal Communications Commission (FCC) needs to assign radio frequencies to seven new radio stations located on the grid in the accompanying figure. Such assignments are based on several considerations including the possibility of creating interference by assigning the same frequency to stations that are too close together. In this simplified situation it is assumed that broadcasts from two stations located within 200 miles of each other will create interference if they broadcast on the same frequency, whereas stations more than 200 miles apart can use the same frequency to broadcast without causing interference with each other. The FCC wants to determine the smallest number of frequencies that can be assigned to the six stations without creating interference.

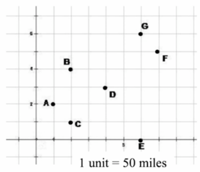


Figure 2

- Student 1 began thinking about the problem by drawing a circle of radius 200 miles centered on each radio station.
- Student 2 began thinking about the problem by drawing line segments to connect pairs of radio stations whenever the two radio stations are within 200 miles of one another.
- Student 3 began thinking about the problem by drawing line segments to connect pairs of radio stations whenever the two radio stations are more than 200 miles from one another.

Which approach seems most promising to you? Use this approach to determine the smallest number of frequencies that can be assigned to the six stations without creating interference. Justify your final answer.

2C The inadequacy of professional development strategies

The importance in educational improvement of professional development for teachers is generally accepted. However fierce their disagreements on other matters, all agree that improvement in the quality of teaching is essential for progress, and that professional development has a key role to play. Every school system has a program (though, when funds are tight, the argument "We want our best teachers in the classroom" is regularly used to sideline it).

There is a wealth of literature on the evaluation of professional development. Classics include Guskey (2000, 2002), Joyce & Showers (1980; 1995) and Loucks-Horsley et al. (1998), Cohen et al (2001). However despite the recommendations from literature, such evaluation is not often designed to provide the kind of feedback needed for the effective design of professional development programs, which requires: well-defined PD designs; observation of the fidelity of their implementation; and detailed observational feedback on teacher classroom behaviour.

Aside from academic researchers, it seems rare for anyone to look for evidence of changes in the behaviour of teachers in the classroom following a professional development program; yet it seems clear that such changes should be the core goal of professional development. Why this mismatch? The dominant approach reflects 'the professional principle' – that teachers take whatever they value from the professional development experience, and that it is not appropriate for one professional to question the judgment or skill of another. This leads to a design approach that seeks 'a civilized discussion between fellow professionals'. Though this approach may work well over a long period for some teachers, the limited range of teaching strategies shown by most teachers suggests that it is inadequate for most.

Those, including ourselves, who have compared teachers' behaviour in their classroom, before and after specific programs, commonly find no observable change. Again why? Professional development programs are usually evaluated by their designers with questionnaires on how far the teachers found the experience valuable – a useful but very different outcome. As ever, feedback has a strong influence on design – programs are designed to be enjoyable for participants, and most do well in this regard. Such a mismatch between the main goals and evaluation criteria exemplifies poor strategic design.

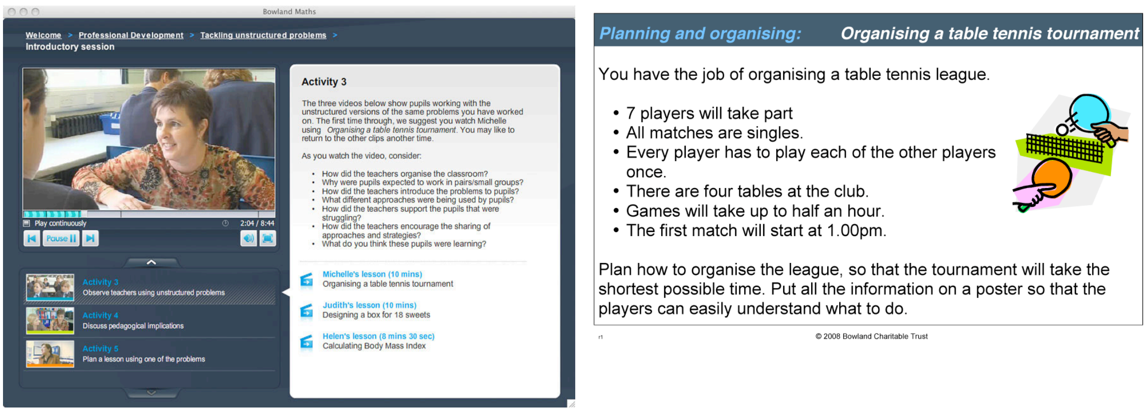
Design challenge

While general pedagogical principles are important in teaching, good teachers also show a wide spectrum of specific high-level skills and teaching strategies. One characteristic, for example, is 'role-shifting' (Phillips et al 1988). Here the students take more responsibility for their own learning and performance, adopting traditional teacher roles (manager, explainer, task-setter). The teacher adopts facilitative roles (adviser, fellow-student, resource), talking less and asking more open and more strategic questions. However the need in this approach to follow students' reasoning and to choose interventions appropriately requires deeper understanding of both pedagogy and the subject. Designing professional development that will enable typical teachers to acquire these new skills is a design challenge.

Over the last few decades, programs that adopt a more skills-focused approach have been developed. The Bowland Maths Professional Development modules (Bowland, 2008) illustrate this - see Figure 8 for an extract from one of the modules. They are based on supporting teachers in trying specific new activities in their classrooms, and reflecting on the experience. General principles are inferred from a sequence of such successful experiences – constructive learning for teachers. Observation shows that teachers make the intended style-shifts, extending their range of classroom strategies and skills – though not surprising, since this is the focus of the design, it is valuable nonetheless. Less clear is how much experience of this kind is needed before teachers carry over these skills into their everyday practice.

This model has been outlined only to show that it is possible to design effective professional development – and that better strategic design and more powerful development methods can both contribute to this.

Figure 8: Extract from *Bowland Maths* Professional Development



(A longer extract is available online)

3. Features of poor strategic design

Strategic design is about ensuring that the product interacts effectively with the system it aims to serve. The examples in Section 2 show lack of understanding of important aspects of the way the system works: that teachers teach to tests; that the difficulty of a substantial task is greater than that of its elements; that discussion of principles is not enough to enable most teachers to acquire new pedagogical skills. When the design ignores such properties, the products may be expected to fail and/or to have undesirable consequences. In this section we explore how this happens, and why it is so common. We start with a few observable surface features, before looking at underlying causes. Such features, illustrated in the above examples, include:

- *Unintended consequences* are a universal feature of poor strategic design. Policy makers assume that their initiatives will achieve their goals without negative side effects. Governments and their educational advisers regularly deplore "teachers teaching to the test". Curriculum specifications are always intended to raise real standards. Professional development programs are only funded because they will forward their goals. Yet such unintended consequences have usually been predicted, by professional designers and some others.
- *Faith in "expert" advice* is an equally common feature of poor strategic design. Policy makers believe that, if they gather together "some of the best minds in the field"[17] , the advice they receive will enable them to achieve their policy goals. That is a natural assumption in their world. Thus tests, standards, and professional development have traditionally been delivered in this way. The reasons it does not work are discussed below.
- *Neglect of "gaming"* Government initiatives are invariably introduced by presenting a set of admirable intentions with which few would quarrel. However, there is little awareness that the rhetoric will be much less influential than specific changes that promise to impact on those involved – "teeth" bite, while talk can and will be ignored. Thus everyone concerned is likely to "game" the system from their own perspective[18] , while Government assumes that the *spirit* of the change will be sustained, however threatening the *detail*.

Underlying causes

Beyond these observable features, what can we say about the causes of poor strategic design? Our understanding of any human system as complex as education will always be incomplete, but there are several common elements in strategic designs that undermine effectiveness.

- *Underestimating the design and development challenge* is an obvious corollary of the mismatch between intentions and outcomes that is shown in the usual failure to achieve the planned positive goals, as well as the unfortunate unintended consequences. Put simply, policy makers and other funders often fail to recognize that there *is* a strategic design problem, that what they want involves more than the straightforward application of the experience of good practitioners. Thus professional design groups regularly face the task of persuading clients to modify the specification so as to increase the chance that the product will achieve their goals.

- *Timescale mismatch* The timescale of politics (a year or two) is shorter than that needed for the design and development of substantial innovations[19], let alone that of implementing real educational improvement (a decade or two). Good design and development takes time, so there is often a conflict with system leadership's wish for quick results.

- *Imbalance between pressure and support* When human beings are asked to do something new, they usually need tools, training and other forms of support to become proficient. For professionals this is particularly important when the change is profound, challenging their beliefs as well as their well-grooved habits of day-by-day practice (see e.g. Fullan 1991). The mathematics teacher trying for the first time to handle discussion in the classroom in a facilitative, non-directive way needs effective help, as does the typical test designer trying, for example, to develop new types of science problems that students will see as relevant to their everyday lives.

  Governments, however, make policy then apply pressure of various kinds to ensure that it is implemented. Underestimating the challenge, less attention is paid to the design, development and provision of support that will enable those involved to implement the policy effectively. The need for support is recognized but what is needed is not realistically evaluated; since effective support tends to be much more expensive than pressure, the need is underestimated. Further, since design and development of effective materials would delay implementation, they hope that "the market will provide" – and, at some level, it will. As for professional development, often an arbitrary sum is allocated for support and its design within this is left to those involved in implementation of the policy.

- *Casual commissioning* Underestimating design and development challenges leads governments and some other funders to underestimate the importance of the commissioning process. On the one hand, they try to specify in too much detail aspects of the design that need to be explored; on the other, often driven by wanting the product urgently, they fail to ensure that the emerging design meets the objectives they had set out. Equally, the selection of the design team is often too casual. A sense of urgency and cost concerns preclude sensible procedures like asking two or three groups to produce outline designs (Burkhardt 2008) along with some trial data on them[20].
- *Unrealistic pace of change* Governments often want to be seen to be "solving the problem", moving swiftly to the end goal – through the introduction of a new curriculum, for example. A one-step approach with a clear end point has advantages. However, the size of the step is limited by how much change those involved can absorb and implement with the support provided through tools and collegial support through networks. Often, the support is less than anticipated, the changes in practice are not made, and the design intentions are undermined.
  The alternative *gradual change* approach defines a direction of change but adjusts

the pace to be digestible. Well-engineered *replacement units* can support profound changes in short bursts. Professional development aims for specific improvements in practice. The main disadvantage of this model is not its speed, which can easily meet the decade timescale of real change, but its lack of glamour – a ‘big bang’ change can inspire both society and (some) professionals.

- *Political need “to be seen to be doing something”* In this media-driven age, where outrage sells better than good news, politicians are constantly bombarded with ‘public’ demands to improve this or that, or to “make sure that this can never happen again”. If examination results improve, the exams are getting easier; if they go down, the education system is to blame. Governments seek to handle this challenge with a string of initiatives [21] to show that they are active in meeting society’s, or is it the media’s, wishes [22].

  This is potentially a source of profound concern since, from this perspective, the success of the initiative in meeting its declared goals is irrelevant. Any comeback will be far in the future when the minister (or even the Government) will have changed; however any initiative turns out, the media will always be able to find something to stir public concern. Most of the politicians I have talked with are not as cynical as this implies but, though they may be keen to do things well, keeping the media at bay is an absolute priority.

This last point generalizes – the priorities of the various key groups in the system that are affected by the product will not be well-aligned. Some will be resistant to change, or simply want a quiet life. Other groups will each have active agendas that may be in conflict. Understanding the system dynamics, and minimizing the impact on the core goals of the design, are the foundation of good strategic design.

Contributions of education professionals

The examples and discussion above may give the impression that bad strategic design is a monopoly of policy makers. However, the education professions are a major contributor. The strategic design of innovations in education, whether for government or other funding agencies, is still usually based on the advice of groups of expert practitioners. Documents are drafted, circulated for comment, and revised, then policies are adopted. But in designing an innovation, such advisers are extrapolating from their own successful experience to the new area in question – and assuming the changes will work well in the hands of other, often less expert, practitioners. Because extrapolation is notoriously unreliable, this craft-based approach can work well for minor changes but, for substantial innovation, it underlies the limited impact and unintended consequences that so often occur.

Typical symptoms of the inadequacy of the input of educational advisers include:

- *They never say that something can’t be done* – while professional educational advisers often criticize government initiatives as “the wrong thing to do”, they rarely say the policy goals cannot be achieved.
- *They never say “We don’t know how to do it”* – probably because, if anyone did,

the client would find someone else, more malleable even if less competent.

- *They never try to define the time and resources that it will take* to design and develop the tools and processes needed to achieve the policy goals; the government or funder takes a decision as to the resources it wants to allocate, to development and to supporting implementation, and the profession accepts that decision, even if the resource allocation guarantees failure.
- *They ignore system realities* of the various kinds described above.

The contrast with research-based professions, like medicine or engineering, is stark. There, research-based methods are used to develop solutions to offer to policy makers, with evidence on their power and limitations. The designers estimate the support needed for successful implementation of the policy, and its costs. When something has not been done before, they say so and estimate the timescale and effort that will be needed to have a good chance of success in that area. So governments don't make policies that are unachievable, or that they cannot afford. (Imagine a research team saying "We'll cure cancer in 5 years with whatever funds you choose to give us" or "We'll have all our energy from nuclear fusion in 10 years".) In this as in other respects, education is more like "alternative medicine" – ever willing to offer a treatment with good faith but with no solid evidence that it works.

The methodology gap

A common feature of the examples in Section 2 is that systematic empirical development through trials, before implementation, would have revealed the sources of failure and might well have suggested improvements in the designs – the standard methodology of systematic development.

How does this happen, for example, in the UK where Government is formally committed[23] to evidence-based policy formation? Indeed, two elements in the standard innovation cycle are now firmly established as part of government policy making. Using medical nomenclature, they are:

- *Diagnosis:* insight-focused research, much of it government commissioned, regularly provides policy makers with diagnostic information on the strengths and weaknesses of current practice in many fields, including education.
- *Phase 3 trials:* pilot field testing of treatments before implementation for evidence on outcomes is Government policy; however, because these strategic decisions are not seen as a design problem by those who make them, the purpose of those at every level who take part in these pilots is to show that the initiative "works", rather than to learn how to improve it. In practice, driven by the short timescale of politics and the need to be seen to be proactive, governments reject only egregious failures.

The key gap in the methodology is *a research-based link between these two elements*, namely: *Design and development* of initiatives using research-based methods.

This is analogous to Phases 1 and 2 of the development of treatments in medicine[24] – the initial small scale Phase 1 explorations leading, in selected successful cases, to their careful systematic development in Phase 2.

Such research-based design and development involves, sequentially:

- R*eview* of research, of craft-based knowledge, and of earlier innovations;
- *Design*, imaginatively exploring a broad range of design possibilities;
- *Development* through an iterative process of feedback from small-scale trials;

sifting out at each stage those candidates and aspects that prove less promising.

Piloting in representative circumstances is the *final* step before large-scale implementation. Its usual role is a summative validation of the initiative, rather than providing formative and developmental feedback. The *prior* phases of research-based development, too-often by-passed in education, are where the product is refined through rich and detailed feedback, its quality and robustness enhanced, and unintended side-effects discovered.

There is a fuller discussion on how to improve the contribution of educational research to practice in (Burkhardt and Schoenfeld 2003) and (Burkhardt 2006) as well as in other contributions to this journal. They point out the many obstacles in the way of useful research that are placed by the current academic value system in education.

There seem to be three main reasons for government resistance to such improvements:

- The lack of awareness of strategic design as an issue that needs as much attention as other aspects of design;
- A reluctance to lose the freedom to make policy decisions based on "common sense" in response to public pressure and/or political opportunity;
- The greater cost and time that professional design implies, modest though the cost is in comparison with the costs of implementation.

All these reflect the belief, widely held in politics and the media, that education is an area where specialized knowledge is needed only for details. "After all, I went through the system and look what it did for me" is a common, usually unspoken, feeling.

4. Successful strategic design: some examples

In this section I outline four initiatives where the strategic design appears to have played a substantial part in their success. This will help to balance the gloomy picture painted so far, showing that effective strategic design *is* possible, and will inform the discussion of principles for strategic design in Section 5. In selecting these examples, I have looked for designs that combine:

- Educational ambition, breaking new ground in the system they serve;
- Some large scale impact (compatible with the goals!);
- Influence on designs that followed; and
- Are in English (with apologies to heterophones).

In each case, there are links and references to more on the materials, including examples [25].

4A Nuffield A-level Physics

This course set out to engage 16-18 year old UK students with the processes of scientific investigation, and to bring some of the major innovations of 20th century physics into school. The origin of this project lay in concerns, common after Sputnik in 1957, about the state of science education and the shortage of scientists. In the absence of a national curriculum specification, this context gave the team freedom to innovate, with success or failure measured by the level of voluntary participation by schools.

In Issue 1 of this journal, Paul Black described the thinking and the effort behind the project, including its strategic design as well as the new and ambitious educational goals and the tactical and technical design moves that were devised to achieve them (Black, 2008). So here I shall be brief, simply bringing together the main strategies.

- The course was developed in collaboration or consultation with the key constituencies, including university physicists, experienced science teachers, both as team members and as participants in trials, schools that would trial the course, equipment manufacturers, publishers, teacher trainers, school district authorities, an examination board, and the funding agency – the Nuffield Foundation.[26]

- A radically new type of A-level examination was developed, reflecting the innovative nature of the course. The unprecedented number of components ranged from a multiple-choice test through more extended examination tasks to a student report on an individual experimental project.
- Since entry to university in England largely depends on the results of A-level examinations, negotiations ensured the wide acceptability of the new examination, in particular for university admissions. The government body charged with oversight of the examinations had also to agree[27].
- The construction of the course was seen as a piece of engineering, a job to be done despite inadequate knowledge of how some of the basic components in the learning process work.
- Two years of trials in a group of schools provided vital feedback, not only for the detailed design, but for acceptability by teachers, and for getting the timing right. These trials did result in some big changes to the original plans.

While the content of the course, which challenged the existing norms for curriculum, pedagogy and examinations, was the core of its success, these strategic elements in the design seem equally essential.

The project had major impact on physics teaching in and beyond the UK. The course and its examination continued for over 25 years, with a related successor now in use. It pushed back the boundaries of what was seen to be possible in school physics, bringing in quantum mechanics and thermodynamics. The project influenced the subsequent development of many more conventional syllabuses and textbooks.

4B Connected Mathematics

This course was designed to improve the teaching and learning of mathematics for US students aged 11 to 14. It was developed through a multi-year project, involving at its peak 12 full-time-equivalent people in the design team. It was funded by the US National Science Foundation, as one of 13 projects that aimed to realize the goals set out in "The NCTM Standards" (NCTM, 1989). Developed by the National Council of Teachers of Mathematics as part of a national concern at the quality of mathematics education, these standards set out learning, teaching and assessment goals for school mathematics across the age range 5 to 18.

Connected Mathematics (CMP), as its name implies, pays particular attention to tactical design issues, including the coherence of, progression in, and connections between the various aspects of mathematics. The curriculum materials build on the authors' decades of experience in prior projects. The contribution in this issue (Lappan, Phillips, 2009) by its lead designers, Glenda Lappan and Elizabeth Phillips[28], sets out the thinking behind their approach and the way they worked.

The strategic design of this and the other NSF-funded mathematics projects followed a standard US model involving: several iterations of planning, design, development, field-testing, and evaluation, followed by publication, marketing, and support – with regular revision to provide new editions.

Some of the design challenges they faced are universal:

- How far can one incorporate changes in the way mathematics is done outside school (using calculators and computers for most routine procedures, for example) and the research findings from the cognitive science and mathematics education on student learning, while remaining acceptable to a society that has a traditional picture of "school math"?
- How far can one demand higher-level teaching skills and still serve current teachers? What support for professional development should one assume?

Other challenges are peculiar to the US, and to this project – for example:

- While in most societies any innovation will face traditionalist counter pressures (and probably should, as a test of its worth), in US mathematics education there is a particularly well-organized, well-funded lobby that attacks any sign of reform.
- Unlike the other three cases in this section, the designers were unable to significantly influence the high-stakes tests that are used for school accountability in all US states, with the usual strong influence on classroom activities. The research indicates that students in schools using CMP perform at least as well on such tests as comparable groups in more traditional curricula but the much higher performance in the extra dimensions of understanding that CMP enables is not assessed or, therefore, publicly recognized.[29]
- Particular attention had to be paid to the listed requirements of the large "adoption states", notably California and Texas, where approval is important for direct impact, influence on other states, and commercial viability through sales. These requirements are often far from coherent, reflecting the diverse wishes of the different groups on the committees that compile them. They always add up to far more than any teacher could teach, let alone students learn[30] – altogether a designers' nightmare that has gotten worse as individual school districts impose "pacing guides" that, week by week, say when each topic should be taught and tested.
- The materials were published and offered for sale in competition with many other curricula, including the four other NSF-funded middle school curricula, some of which have strong features, as well as the traditional curricula that have long dominated the market. (These might be seen as 'comparison groups'.)

In spite of these formidable challenges, CMP has achieved major impact on US schools. It has a substantial share of the market and is central to any discussion of middle school mathematics education. What are the factors behind this success?

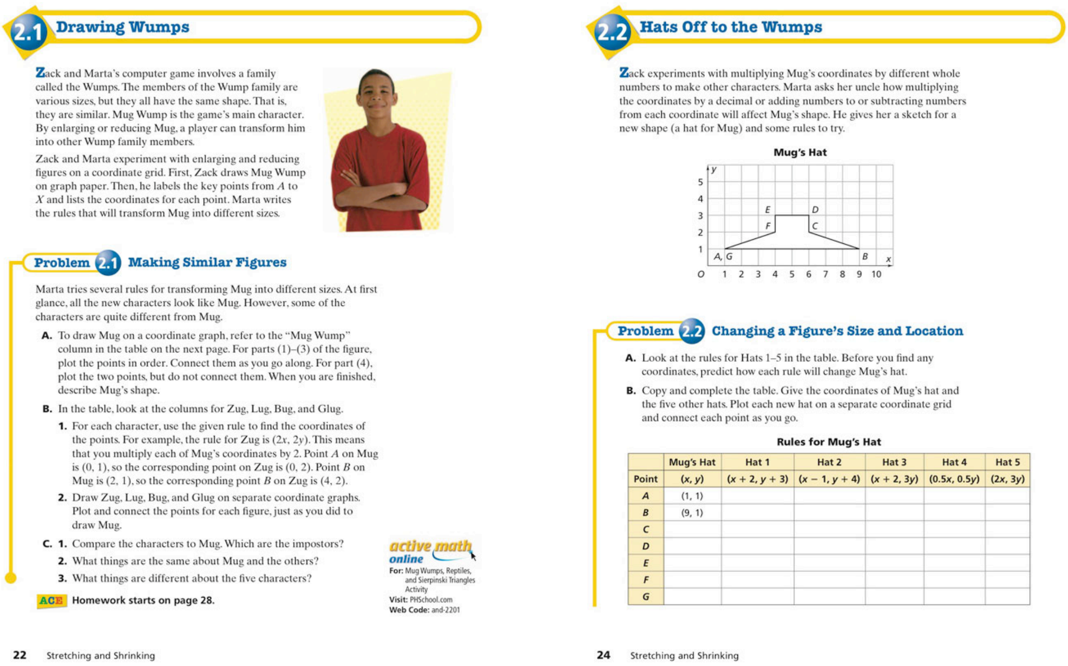
- *Inherent quality of the material* All authors will say that this, the tactical and technical design, is the key to success. One would like this to be true, and the quality of CMP is widely acknowledged. On the other hand, the traditional curricula that have dominated the market, and still have a large share, succeed despite their only obvious virtue being familiarity to the customer and client groups[31]. "That's the proper way to teach math, like when I was at school."

- *Quality of the design team* The success of the *Connected Mathematics* curriculum, which is written for both teachers and students, reflects the diverse talents of the design team. The team consisted of authors, graduate students, graphics designers, teacher collaborators, researchers, and an advisory board made up of mathematicians, mathematics educators, teachers, administrators, and parents.In addition, consultants from the sciences, engineering, reading, English language learners, and special education provided valuable insights for specific aspects of the curriculum.
- *Understanding and growing a "niche market"* The authors have long been at the heart of the main organizations in US mathematics education, not only NCTM but NCSM, the smaller organization of "math supervisors" in school systems who strongly influence the choices of materials. This has given them a deep understanding of the needs and constraints perceived by these key constituencies. This has informed the design of CMP. Those who were looking for real change, long advocated within this community, found a workable curriculum that met their ambitions.

- *Specific support for meeting strategic challenges* The project recognized the challenges that implementation presents and offered specific guidance and support. Figure 9 shows examples of this in CM materials.
- *National evaluation* Driven by the "math wars" controversy, the Bush administration commissioned an evaluation of the available curricula[32]. An expert group (not, on this occasion, pre-selected to produce "the right result") rated *Connected Mathematics* as exemplary.

The future will show how far the continuing counter-campaign will succeed, or whether CMP will provide the new base from which further advances can be built – for example, in the fuller integration of IT, and the delivery of functional mathematical literacy.

Figure 9: Example from *Connected Math: Stretching and Shrinking* (sample)
(A longer extract is available online)



2.1 Drawing Wumps

Zack and Marta's computer game involves a family called the Wumps. The members of the Wump family are various sizes, but they all have the same shape. That is, they are similar. Mug Wump is the game's main character. By enlarging or reducing Mug, a player can transform him into other Wump family members.

Zack and Marta experiment with enlarging and reducing figures on a coordinate grid. First, Zack draws Mug Wump on graph paper. Then, he labels the key points from *A* to *X* and lists the coordinates for each point. Marta writes the rules that will transform Mug into different sizes.

Problem 2.1 Making Similar Figures

Marta tries several rules for transforming Mug into different sizes. At first glance, all the new characters look like Mug. However, some of the characters are quite different from Mug.

A. To draw Mug on a coordinate graph, refer to the "Mug Wump" column in the table on the next page. For parts (1)–(3) of the figure, plot the points in order. Connect them as you go along. For part (4), plot the two points, but do not connect them. When you are finished, describe Mug's shape.

B. In the table, look at the columns for Zug, Lug, Bug, and Glug.

1. For each character, use the given rule to find the coordinates of the points. For example, the rule for Zug is $(2x, 2y)$. This means that you multiply each of Mug's coordinates by 2. Point *A* on Mug is (0, 1), so the corresponding point on Zug is (0, 2). Point *B* on Mug is (2, 1), so the corresponding point *B* on Zug is (4, 2).
2. Draw Zug, Lug, Bug, and Glug on separate coordinate graphs. Plot and connect the points for each figure, just as you did to draw Mug.

C. 1. Compare the characters to Mug. Which are the impostors?
2. What things are the same about Mug and the others?
3. What things are different about the five characters?

ACE Homework starts on page 28.

active math online
For: Mug Wumps, Reptiles, and Sierpinski Triangles Activity
Visit: PHSchool.com
Web Code: and-2201

22 Stretching and Shrinking

2.2 Hats Off to the Wumps

Zack experiments with multiplying Mug's coordinates by different whole numbers to make other characters. Marta asks her uncle how multiplying the coordinates by a decimal or adding numbers to or subtracting numbers from each coordinate will affect Mug's shape. He gives her a sketch for a new shape (a hat for Mug) and some rules to try.

Problem 2.2 Changing a Figure's Size and Location

A. Look at the rules for Hats 1–5 in the table. Before you find any coordinates, predict how each rule will change Mug's hat.

B. Copy and complete the table. Give the coordinates of Mug's hat and the five other hats. Plot each new hat on a separate coordinate grid and connect each point as you go.

Rules for Mug's Hat

	Mug's Hat	Hat 1	Hat 2	Hat 3	Hat 4	Hat 5
Point	(x, y)	$(x + 2, y + 3)$	$(x - 1, y + 4)$	$(x + 2, 3y)$	$(0.5x, 0.5y)$	$(2x, 3y)$
A	(1, 1)					
B	(9, 1)					
C						
D						
E						
F						
G						

24 Stretching and Shrinking

4C VCE Mathematics

In the late 1980s, the Victoria Certificate of Education was introduced to all Victorian schools as a single pathway for all students to complete secondary school and, at the same time, as a way in which universities could select students for particular courses of study. The VCE was designed as a course of study to be taken over two years in a range of subjects, constructed according to the same set of principles and accredited by a single authority representing government and other key stakeholders.

Assessment within the VCE would be a mix of school-based assessments and end-of-year examinations. Under the Mathematics Study Design, the course had to provide time for teaching and learning in:

- The development of standard skills and applications;
- Problem solving, applications and modeling (hereafter called problem solving); and
- Mathematical investigations (hereafter called projects);

Students had to demonstrate that they had worked on all these 'work requirements' in both Years 11 and 12. For all final year (Year 12) mathematics courses, the assessment balance was set at 33% for school assessed coursework and 67% for end-of-year externally set and externally graded examinations. Students' work was assessed by their teachers and the results were moderated by groups of teachers from nearby schools. Here we report only on those changes related to the introduction of problem solving and projects.

VCE mathematics took a fresh look at the range of types of performance that are important in mathematics, and developed ways to assess the expanded range in a high-stakes assessment. The design of the *problem solving and modeling* coursework broke new ground in many ways. While the timed examinations were based on standard task types, the VCE included the following innovative features:

- Mandating that, as part of the assessment, students tackle non-routine problems and mathematical investigations ('projects') in both pure mathematical and real world contexts;
- Providing substantial time for these tasks, both in class and at home, with strict protocols for teachers to authenticate that the work done outside examination conditions had really been done by the student;
- Providing each year new "starting points" and "themes" for problem solving and modeling tasks and projects that were compulsory for students to work on – on the one hand these gave students the opportunity to define their own specific problems and solution paths and, on the other hand, ensured some mathematical depth in the topics involved;
- Developing criteria for scoring students reports on their problem solving and projects, along with systems for teachers to moderate results across schools; and
- Designing a test to authenticate coursework.

In the early years of the new examination, students had to undertake a 20-hour mathematical project over 4 weeks, and an 8-hour problem solving task over 2 weeks in each mathematics subject[33]; this was later changed because of workload so that there was only one of these tasks for each subject.

The genesis of this innovation involved people who were at the forefront of Australian developments in mathematics education. Ross Turner and later Max Stephens managed the design and implementation and smoothed its passage into reality, always a challenge for innovative high-stakes assessment. (VCE results are a key factor in university entrance decisions.) Susie Groves and Kaye Stacey[34] had pioneered the introduction of problem solving into teacher education at Burwood College, now in Deakin University, with "The Burwood Box" and associated teaching materials for schools (Stacey and Groves, 1985). Both were seconded to the examination board to develop the very substantial written support materials, which explained the new processes of problem solving and modeling to teachers. When concerns from universities about standards and authentication demanded revisions, Peter Stacey and Barry McCrae played a leading role in the re-design process, including the authenticating test.

For about the first decade of VCE Mathematics, the assessment tasks were developed each year by groups including university mathematicians, mathematics educators and practising teachers, and provided to schools by the central assessing authority. They showed teachers the activities that were important for students to engage in, and provided topics that contained substantial mathematical content related to the course material. Sample scripts at each grade level, marked and annotated, were supplied to ensure consistent marking by teachers and assessment supervisors.

Figure 10 shows a brief example of a state-provided theme for an “Investigative project” on mathematical modeling and rates of change, with one of the starting points that students could choose for the 20-hour project for the main calculus and functions subject. Figure 11 shows the complete task and instructions for another theme, Maxima and minima. Note that the starting points have a structured part (a), but encourage students to work independently and to follow their own paths in part (b). Students would work in class to begin the project then, over 4 weeks in class and at home, would carry out the investigation and prepare the report. Students would report regularly to their teacher to provide evidence that they were working on the projects themselves, and finally submit a report of about 10 pages for assessment. Figure 12 shows the criteria for teachers to use in assessing student work including an assessment checklist and grade descriptors.

Some teachers and students found the experience of the project stressful, and indeed some misunderstood the nature of mathematical investigation so wasted time preparing extraordinarily visually attractive reports with minimal mathematical content. However, for many teachers and students, these activities provided an unsurpassed mathematical experience. Stacey (1995, p 66) quotes one very experienced teacher as saying: “I have never seen such intense, creative and cooperative work in mathematics. In class, there was a great deal of discussion, yet they were all working on their own problems.”

Figure 10: VCE “Investigative Project” theme example
A longer extract is available online

Theme: Mathematical modelling and rates of change

Your project must be based on, or incorporate, mathematical modelling and rates of change. You should use mathematics appropriate to the focus of Change and Approximation and Extensions (Change and Approximation). You are encouraged to show initiative and be independent in carrying out your project.

General Advice

In this investigation, you are required to examine the way in which different variables are connected in a mathematical model of a real situation. In particular you are asked to examine the connections between the rates of change of different variables in the model.

As a simple example, consider what happens when you pour liquid into a container. Obviously, the faster you pour, the faster the height of liquid will change. The height of the liquid is a function of the volume of liquid poured into the container. The rate of change of height depends upon the rate of pouring and the shape of the container. For example, in a cylindrical container, the rate of change of height is directly proportional to the rate of pouring, but if the container is of a more complex shape (eg a conical flask) this relationship will be more complicated.

In exploring these connections, it may be appropriate, in some cases, to perform simple experiments to approximate what happens, but it is essential that you develop a mathematical model which represents the situation and that you use mathematical techniques to establish results and connections. (It may, of course, be useful to compare mathematical with experimental results, but **experiment alone is not sufficient.**)

Where you use a mathematical function derived from an experiment, or data, you are **not** required to provide a theoretical justification for your choice of function.

Mathematical modelling: In developing a mathematical model to represent any of the situations given in the starting points, the following steps may be helpful.

1. Make assumptions which simplify a real situation. In the example above, you might ignore evaporation or assume a container to be a perfect cylinder. Consider the likely effects of these assumptions on your results.
2. Identify the key variables in your model. If there are too many variables to deal with initially, assume some of them to be constant so that you are left with a problem which you can work on mathematically. The choice of constants will, in general, be left up to you, but you should give some consideration to making realistic choices.
3. If you feel that you have fully explored a model with the assumptions you have made, it may be appropriate to change one, or more, of your assumptions. In most cases, the Starting Points suggest how this may be done.
4. At each stage, you should attempt to evaluate what your results are telling you. Look carefully at the domain in which your results might be valid and consider what might happen with extreme values of your variables. Consider the practical significance of your results.

Starting Points

You must investigate ONE of the following Starting Points. You must confirm with your teacher your choice of Starting Point (preferably by the end of the first week of the designated period).

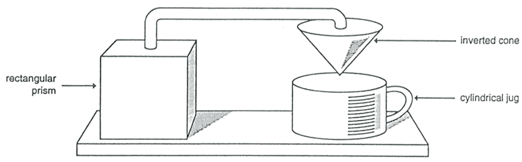
1. Filter coffee-makers

In this starting point you should examine the connections between the flow of liquid in the different parts of a filter coffee-maker. You are to explore the relationships between the height and volume of liquid and how they vary with time.

A filter coffee system consists of three connected containers. Water is heated in a container, pumped through to the filter which contains the filter-paper and coffee-grounds and then drips into the coffee jug.

a. Working with a simple model

You should assume that water is initially drawn from a container which can be regarded as a **rectangular prism** and is pumped to a filter **which has the shape of an inverted right circular cone.** It then drips into a **cylindrical** coffee jug.



i To begin, assume that water is pumped into the filter at a constant rate, and that the filter is blocked. That is, we are only looking at the first stage of transfer of liquid. Using a rate of flow which would allow a litre of water to pass into the filter in 15–20 minutes and by assuming that the filter is large enough to hold this amount of water, find the rate at which the volume and the height of water in the filter change with time. Sketch graphs of the rates of change which you have found.

ii From these formulae for rates of change, you should now be able to find, by anti-differentiation or otherwise, formulae for volume and height in terms of time.

iii Now look at the second stage of transfer of liquid. Imagine that the filter is full of liquid. Allow this liquid from the filter to flow into the jug without any more water being pumped into the filter. Assume that the rate of flow is proportional to $\sqrt{h}$, where h is the height of liquid in the filter at time t. Find rate of change formulae, with respect to time for height and volume of both the liquid remaining in the filter and the liquid filling the jug.

iv From the various formulae for rates of change, you should now be able to find, by anti-differentiation, formulae for the volume and height of liquid in the jug in terms of time. Alternatively, sketch graphs of the rate of change functions to describe the behaviour of the system.

b. Extending the model

Here are a number of suggestions for extending the simple model given above. You are required to pursue only **one** of these **or** a similar alternative.

i Consider the whole system operating simultaneously. In particular this means looking at the height of liquid in the filter when water is pouring in and dripping out of the filter.

ii You may wish to consider jugs of different shapes and/or the problem of where to place measurements showing cup gradations on the side of the jug. Which cup is filled the quickest?

iii The assumption that water flows into the filter at a constant rate is questionable. In practice it starts slowly, builds up to a maximum and then decreases. Find a simple function with these properties over the domain under consideration and explore the effect of using this rate of flow on the rate of change of height and volume of water in the filter.

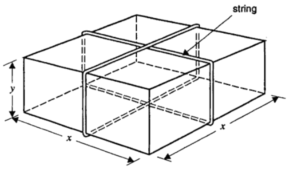
Figure 11: VCE *Maxima and Minima* investigation tasks
A longer extract is available online

3. How long is a piece of string?

This starting point explores the length of string or straps required to secure parcels or loads of varying dimensions. You will be concerned with choosing the dimensions of a load (which can be regarded as a rectangular box) of given volume, so that the length of string or straps required is a minimum.

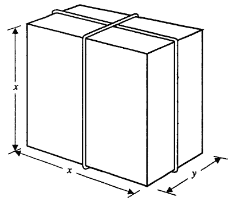
a.

i. A parcel is to be tied with a single length of string in the normal way (see diagram below).

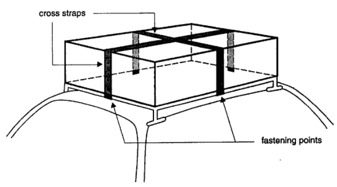


Assume that the parcel is a cuboid with square cross-section and with a fixed volume. Find an expression for the length of the string and determine the dimensions for the parcel which minimise the length of string required. Sketch a graph of the function obtained for the length of the string and comment on the physical implications of this graph. What difference does it make if the extra string required to tie the knot is included?

ii. If the parcel were to be tied on its side, that is, so that the strings do not cross on the square face (see diagram below), how does this change the solution?



b. i. Now consider the problem of strapping a large package to the roof-rack of a car (see diagram below).



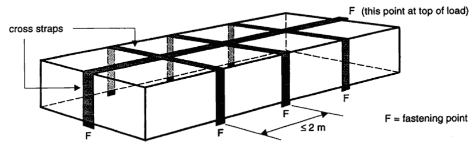
Assume that the width of the package is fixed (that is, determined by the width of the car or the rack) and that the straps do not go under the package. The length and height of the package are not fixed. Find the shortest total length of cross-straps required to secure the load, which is a cuboid of fixed volume. You may find it useful to choose suitable values for width and volume. Comment on the feasibility of your solution.

ii. If two cross-straps instead of one were required to go across the width of the car, how would this affect the best dimensions of the load?

iii. Repeat parts **i.** and **ii.** with different values for width and volume, or try to generalise your solution in some way. Comment on the practical implications of your solutions.

c. Extension

i. Consider the general problem of securing a large load to a truck. The load is of given width, but variable length and height.



Imagine you are responsible for securing a load to a truck as in the diagram above. (Straps do not go under the load.) The rule is that the cross-straps to be used for fastening the load must be no more than two metres apart along its length. Your task is to experiment with different values for the volume. In each case, find the dimensions of the load which requires the least length of straps. What is the relationship between volume, number of cross-straps needed and height of the load?

ii. Vary the rules for strapping in a way you consider realistic. How, for example, would you deal with an unusually wide load? For your particular variation, find relationships between volume, number of cross-straps required, height of load and total length of cross-straps.

Figure 12: VCE Assessment Criteria
A longer extract is available online

CAT 1: Investigative Project

ASSESSMENT CHECKLIST

VCAB School Number: ☐☐☐☐☐ Teacher attestation completed: Yes ☐ No ☐

Candidate Number: ☐☐☐☐☐☐☐ Teacher name:

Conducting the investigation	Not shown	Low	Med	High
identifying important information	☐	☐	☐	☐
collecting appropriate information	☐	☐	☐	☐
analysing information	☐	☐	☐	☐
interpreting and critically evaluating results	☐	☐	☐	☐
working logically	☐	☐	☐	☐
breadth or depth of investigation	☐	☐	☐	☐

Mathematical content	Not shown	Low	Med	High
mathematical formulation or interpretation of problem, situation or issue	☐	☐	☐	☐
relevance of mathematics used	☐	☐	☐	☐
level of mathematics used	☐	☐	☐	☐
use of mathematical language, symbols and conventions	☐	☐	☐	☐
understanding, interpretation and evaluation of mathematics used	☐	☐	☐	☐
accurate use of mathematics	☐	☐	☐	☐

Communication	Not shown	Low	Med	High
clarity of aims of project	☐	☐	☐	☐
relating topic to theme	☐	☐	☐	☐
defining mathematical symbols used	☐	☐	☐	☐
account of investigation and conclusions	☐	☐	☐	☐
evaluation of conclusions	☐	☐	☐	☐
organisation of material	☐	☐	☐	☐

Grade descriptors for CAT 1: Investigative Project

A. Clearly defined the investigation and evaluated the conclusions. Demonstrated high level skills of organisation, analysis and evaluation in the conduct of the investigation. Used high level mathematics appropriate to the task with accuracy. Communicated the results succinctly in the specified project report format.

B. Clearly defined the investigation. Demonstrated skills of organisation, analysis and evaluation in the conduct of the investigation. Used mathematics appropriate to the task with accuracy. Communicated the results clearly in the specified project report format.

C. Defined the investigation. Demonstrated some facility in the collection and analysis of appropriate information. Used mathematics appropriate to the task. Communicated the results in the specified project report format.

D. Defined the investigation. Identified and collected appropriate information. Used mathematics relevant to the task. Completed the report in the specified format.

E. Stated a project topic relevant to the theme. Identified basic information. Used mathematics relevant to the task. Completed the report in the specified format.

Figure 13 shows three examples of the “problem solving task”: *Oil pipelines; Through the fog; Rational points on curves*. Designed for 8 hours work, in and out of class, these tasks recognized that students require time to conduct substantial mathematical problem solving of a non-routine nature. Since these tasks are more structured than the projects, providing a set of non-routine questions for students to tackle, they give students less opportunity to follow their own paths. Public concern grew in the first years that some students were getting unauthorized help – with (unsubstantiated) rumours of “buying solutions in the market”. Protocols for teachers to monitor each student’s progress worked well in many schools, but in the high stakes environment, suspicion about cheating lurked.

This problem was solved by the introduction of an interesting innovation – a short timed test (again centrally set) on the main mathematical ideas involved in the solution of the problem, given to students after their reports were handed in. Figure 14 shows the test for students who had tackled *Through the fog*. Students whose performance on the test did not match their performance on the 8-hour task were called for interview by teachers and principals, where they were given another opportunity to demonstrate their understanding of the mathematics in their reports. This process worked very well, and restored public confidence in the assessment (McCrae, 1995; Stephens & McCrae, 1995). Teachers were supported in the assessment challenges through published support materials, including task-specific criteria and mark schemes (Figure 15 shows the criteria for *Oil Pipelines*) illustrated with student work.

Figure 13: VCE *problem solving* tasks
A longer extract is available online

Specific instructions to students

You must choose only **one** of the three problems below. You are expected to attempt all parts of the problem which you choose. The task is designed as a series of questions, the purpose of which is to aid students to explore the problem; however, it is not essential that the report consist of a series of separate answers to these questions.

Note carefully the suggested mathematical techniques for each problem. These should provide some hints as to how to proceed and will also indicate the scope, **but not necessarily the context,** of the test which follows upon the completion of this report.

The use of technology is encouraged. However, in using a computer spreadsheet, for example, you must ensure that the appropriate formulation is included and all variables and units are correctly defined. All graphs, including computer generated graphs, must be appropriately scaled and correctly labelled so that key features of the graph are evident.

Problem 1 – Oil pipelines

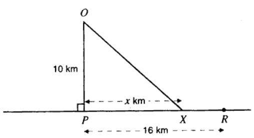
The mathematical techniques which might be required for this task include:

calculus – differentiation using chain rule, minimisation
functions – domain, range, sketch graphs
Pythagoras and elementary trigonometry
algebra – quadratic equations

While any other prescribed methods are acceptable, the above are considered particularly appropriate, and may feature in the test which will follow this task.

An oil platform is situated at a point *0*, at sea, 10 km from the nearest point *P* on a stretch of straight coastline. An oil refinery is at a point *R*, 16 km along the coast from the point *P* (see Figure 1 below). It is necessary to lay a pipeline from the platform to the refinery. This is to consist of straight line sections. Laying pipeline underwater is $\frac{4}{3}$ times as expensive as laying pipeline on land. This ratio is known as the *relative cost*.

Figure 1



Question 1

If the pipeline reaches the coast at a point *X*, distance *x* km from *P* in the direction of *R*, find an expression for the cost *C* (in dollars) of laying the pipeline in terms of *x*, given that laying pipeline on land has a fixed cost of *A* dollars per kilometre.

Problem 1 – continued

Question 2

Find the least expensive route for laying the required pipeline.

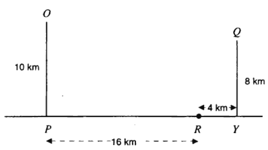
Question 3

Compare the cost of the solution you have found with the cost of a pipeline directly from *O* to *R* and with a pipeline from *O* to *P* to *R*.

Question 4

If the cost of the underwater pipeline changes in relation to the cost of the coastal pipeline, explore how the least expensive route changes. In particular, see if there is a value of *k* (the relative cost) at which the direct *OR* route becomes the cheapest solution, or one for which the route from *O* to *P* to *R* is the least expensive.

Figure 2



Question 5

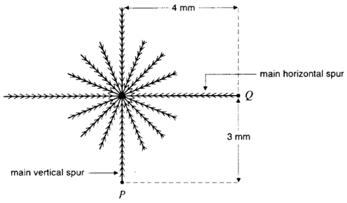
Another platform is to be located at a point *Q*, 8 km from shore on the other side of *R*, opposite a point *Y*, 4 km from *R* (see Figure 2). With the relative cost of $\frac{4}{3}$ as before, explore whether there are any cost advantages in combining pipelines, stating any assumptions you have made.

Figure 14: VCE Post-investigation test for *Through The Fog* investigation tasks A longer extract is available online

Test 2 – Crystal growth patterns

This test is to be attempted by students who completed Problem 2 – Through the fog as their problem-solving task.

A research chemist is studying the way in which crystals of a certain chemical compound form. This diagram shows the features of a *standard seed* crystal of the substance.



The chemist decides to define the size of the crystal as the distance between the end points of the two main spurs, marked P and Q in the diagram. Thus the size in millimetres of the standard seed crystal shown above is

$$\sqrt{3^2 + 4^2} = \sqrt{25} = 5.$$

The chemist finds that the main horizontal spur grows more quickly than the main vertical spur. Under optimum conditions, the main horizontal spur grows at a rate of 3 mm per hour and the main vertical spur grows at a rate of 2 mm per hour.

In the following calculations, give your answers rounded to three decimal places of a millimetre or an hour.

Question 1

Find the size of a crystal, grown from a standard seed crystal, after two hours growth under optimum conditions.

2 marks

Question 2

A function s relates the time, t hours, to the size of the crystal grown from a standard seed crystal under optimum conditions.

a. Show that $s(t) = \sqrt{13t^2 + 36t + 25}$

b. If the domain of s is $[0, 5]$, find its range.

3 marks

Question 3

Find how many hours it takes for a crystal grown from a standard seed crystal to reach a size of 20 mm under optimum conditions.

4 marks

Question 4

Define a function which can be used to find the overall growth rate (that is, the rate of change of size) of a standard seed crystal in mm per hour under optimum conditions. Use it to find the rate of growth of a crystal after three hours.

3 marks

Figure 15: VCE Assessment Criteria for post-investigation test

Criterion Specific Advice

Problem 1 – Oil pipelines

Criterion 1
Initial assumptions might include: working only in two dimensions, that is, depth of the sea will not affect length of pipeline or cost; connections do not add to the overall cost, etc. Clear diagrams and other methods are essential for illustrating variables, constraints and assumptions. Assumptions relating to Question 5 must be made clear. See solution notes for examples.

Criterion 2
Solution notes should be used as a guide. Reasons for the choice of particular techniques should be clearly stated. Terms need to be defined wherever spreadsheets are used, and their formulation explained.

Criterion 3
Opportunities exist to use graphs to represent, for example, the total cost function in terms of the distance x, for $0 \le x \le 6$. In Question 4, opportunities exist for analysing the relationship between x and k for minimum values of the total cost (see solution notes). Graphs or other techniques should be used to indicate whether a point is a local maximum or minimum.

Criterion 4
All diagrams should be clearly labelled. Variables should be chosen appropriately and be clearly defined. Graphs should have a title with each axis correctly positioned and labelled. All graphs, including computer generated graphs, must be appropriately scaled and correctly labelled so that key features of the graph are evident. Asymptotes should be clearly identified. Equations should be formulated correctly. Headings of columns on spreadsheets and tables, where used, should have clearly described.

Criterion 5
Use solution notes as guide to the correct and accurate use of mathematics.

Criterion 6
Solution notes should be used as a guide to the mathematical qualities necessary for a simple, efficient and effective solution. The notes also indicate where insight and elegance should be rewarded.

Criterion 7
Answers to each question need to be evaluated and interpreted in terms of assumptions a student has made. Results should be placed in the context of the problem being solved and any simplifying assumptions made along the way. This will be especially relevant to Questions 5 and 6. Students should use and comment upon a sensible degree of accuracy in results. Unrealistic degrees of accuracy should be avoided.

Criterion 8
Question 5 provides opportunities to link previous results together. Simplifying assumptions should be made clear. Limitation attaching to results will need to take account of constraints and assumptions made in earlier parts of the problem. These should be evident in the way conclusions are formulated.

Criterion 9
Is the report well organised? Does it clearly indicate that the student understood the problem? Does it address all stages of the problem? Are the solution processes clearly evident and the conclusions carefully stated?

Criterion 10
All stages of the problem need to be thoroughly investigated. A thorough analysis should be given for Question 5 and all conclusions should be justified in clear mathematical terms.

Problem 2 – Through the fog

Criterion 1
Initial assumptions might include reference to the absence of cross currents, wind, etc. The focus is simply on the motion of the two boats each conceived of as a point moving in a two dimensional plane. In Question 8, students should recognise realistic constraints on the values of y and z, for example, both y and z never become zero. A realistic domain for t is also assumed.

Criterion 2
Solution notes should be used as a guide. However, alternative formulations might include use of vectors, or coordinate systems with appropriate sign conventions.

Criterion 3
Opportunities exist to use graphs in Question 3 and in Question 5 in order to analyse the behaviour of the functions under consideration. Graphs or other techniques should be used to determine whether a point is a local maximum or local minimum.

Criteria 4, 5, 6
See general comments for Problem 1.

Criterion 7
See general comments for Problem 1. Interpretation and evaluation will be especially important for Questions 6 and 8. See solution notes for examples.

Criterion 8
Questions 6, 7 and 8 provide opportunities to link previous results together. Simplifying assumptions should be made clear, for example, that the boats continue to move in a straight line, and that at no stage are they assumed to reverse their direction of motion.

What were the features of strategic design behind this success?

- *Changing a high-stakes examination* The research study described below (Barnes, Clarke and Stephens, 2000) shows how the change in the high-stakes examination led to corresponding changes in classroom practice throughout secondary school.
- *A fine balance of ambition and realism* The changes sought and achieved were very ambitious, reflecting a world-wide consensus on best practice in school mathematics in a form that was (and is) rarely found.
- *Development from feedback* The curriculum and assessment authority was prepared to fine-tune the design to make it work.
- *A consensus for change in the system* The changes in VCE were mandated by the

system as a whole; this design team used the opportunity provided by a climate in which "the status quo was not acceptable".

- *Tests as a monitoring device* Assessing and authenticating extended pieces of student work is always a challenge. Tests have been used in various ways as part of this process; the use here as a monitoring device was both original and effective in restoring public confidence.

In an associated research study, Barnes, Clarke and Stephens (2000) looked at changes in what happened in school classrooms following this change in high-stakes assessment. This compared classrooms in Victoria with those in New South Wales where, though the rhetoric promoting problem solving was similar, there had not been corresponding reforms in assessment. They found that problem solving activities involving mathematical tasks of the kind introduced into the tests were introduced into classrooms, not only in the final year but throughout the secondary schools involved – though, in the lower grades, perhaps more in form than in substance. David Clarke wrote:

> "Most striking in this analysis, was the evidence in Victoria of the 'ripple effect' (Clarke & Stephens, 1996), whereby the language and format of teacher-devised assessment tasks employed in grades 7 to 10 in Victorian schools echoed their officially mandated correlates in the 12th grade VCE to an extraordinary level of detail. "

The classroom visibility of problem solving activities and assessment emerged as the key difference between the two states.

Because it tracked *changes*, this study is important in providing evidence of a *causal* connection between task types in high-stakes assessment and activities in the implemented curriculum – not simply the well-known similarity of the two (see section 2A).

As often happens, the success of this assessment model was ultimately undermined by outside events – problems in subjects other than mathematics caused the curriculum and assessment authority to remove any restrictions on the *type* of school-based assessment. Gradually, schools decided it was easier to give assessment that mimicked the remaining examinations, and so the experience for students of engaging in substantial problem solving and investigations gradually withered.

Were there weaknesses in the strategic design? Whenever the school-assessed component was strongly guided by official requirements and material for substantial investigations was supplied, it went well in the refined system. But when both formal requirements and support were withdrawn, it was seen as too challenging. This suggests:

- *Task design is critical* The system-level decision to give teachers freedom in the choice of assessment tasks again was a crucial error in strategic design. In other systems this move has often been justified as "giving teachers and students freedom for creative work" but, in a high-stakes assessment, it is no surprise that they give security and predictability a higher priority. Control over, at least, the range of task-types used and how frequently the tasks must be changed is central to the assessment of non-routine problem solving and investigation.

- *Performance goals should be spread across task types* The designers' decision to focus the formal examinations entirely on "facts and skills" and "analysis" (slightly longer but still routine questions) meant that problem solving and investigations, confined to the school-assessed component, were vulnerable to the above change. This decision seemed to make sense, given the time appropriate for substantial investigative activities. (The case study in section 4D, below, shows that such things can be assessed to some extent in timed examinations.)
- *Support materials are essential* Exemplar problems and projects were important in supporting teachers and students in this new kind of work; however, they were only developed for Year 12, whereas teachers of earlier years could have benefited similarly, raising levels of performance throughout the schools.

A notable feature of this innovation, in comparison with the others in this Section, is the fluidity of design control. Key decisions were taken by committees with changing chairs and membership; the coherence of approach that was maintained over a decade perhaps reflects that of the mathematics education community in Victoria at that time.

The initiative has had effects in other Australian states that persist, with some increasing emphasis on real problem solving nationwide (see e.g. Curtis & Denton, 2003).

4D Testing Strategic Skills – "The Box Model"

This initiative, developed in the 1980s by the Shell Centre with the largest UK examination board (Joint Matriculation Board, JMB), brought together in a single package (presented as a box of materials - see Figure 16):

- A new type of task for a high-stakes mathematics examination – with five task exemplars, designed to show the variety to be expected in the 'live' examination, with scoring guidance and examples of student work;
- Teaching materials for three weeks' teaching, developed to enable typical teachers to prepare their students for this type of task; and
- Materials to support related in-school do-it-yourself professional development.

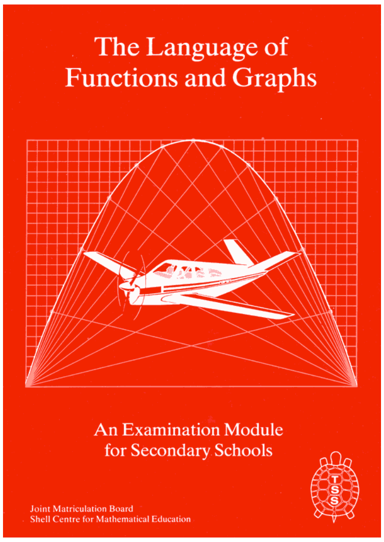
An unusual strategic design feature, compared to the examples outlined above, was the *gradual change* model that was adopted. One new task type was introduced each year, representing:

- One question on the examination;
- 5% of the two-year mathematics syllabus; and
- About three weeks teaching.

Care was taken to remove from the syllabus some topics that took a comparable amount of classroom time. This approach proved popular with teachers. They enjoyed the three weeks of new teaching, pedagogically challenging but well-supported; they were equally glad to get back to more familiar ground for a while thereafter. They looked forward to the next package.

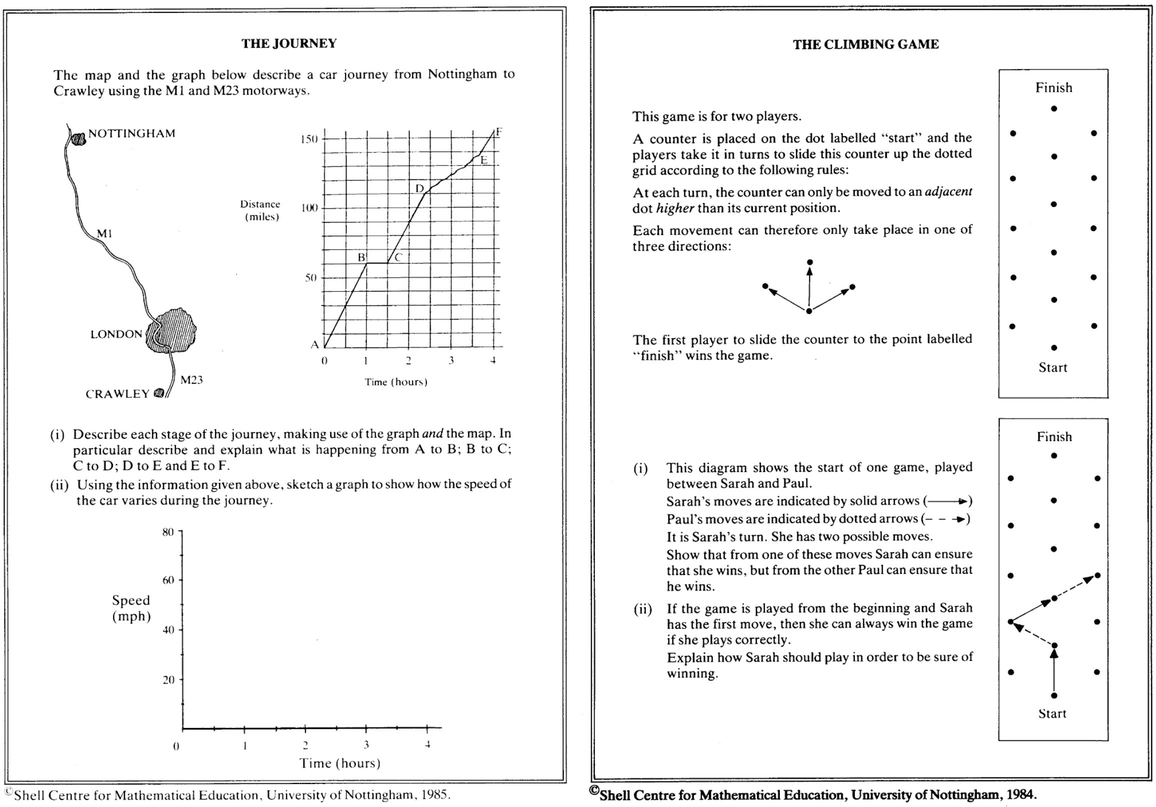
The first year's change was the introduction of 15-minute tasks that assess non-routine problem solving in pure mathematics. The materials, published as *Problems with Patterns and Numbers* (PPN, Shell Centre, 1984), were bought by most of the schools that took the Board's O-level examination for age 16 students. The following year, *The Language of Functions and Graphs*[35] (LFG, Swan, 1986) introduced the modeling of real world situations with Cartesian graphs, and with algebra – graph interpretation, model critique and formulation are all included. See Figure 17 for exemplar tasks from both boxes.

Figure 16: Testing Strategic Skills - summary of contents
(A longer extract is available online)



EXPANDED CONTENTS

Figure 17: Exemplar tasks from *Testing Strategic Skills* (A longer extract is available online)


THE JOURNEY

The map and the graph below describe a car journey from Nottingham to Crawley using the M1 and M23 motorways.

(i) Describe each stage of the journey, making use of the graph *and* the map. In particular describe and explain what is happening from A to B; B to C; C to D; D to E and E to F.

(ii) Using the information given above, sketch a graph to show how the speed of the car varies during the journey.

THE CLIMBING GAME

This game is for two players.

A counter is placed on the dot labelled "start" and the players take it in turns to slide this counter up the dotted grid according to the following rules:

At each turn, the counter can only be moved to an *adjacent* dot *higher* than its current position.

Each movement can therefore only take place in one of three directions:

The first player to slide the counter to the point labelled "finish" wins the game.

(i) This diagram shows the start of one game, played between Sarah and Paul.
Sarah's moves are indicated by solid arrows (——►)
Paul's moves are indicated by dotted arrows (– – –►)
It is Sarah's turn. She has two possible moves.
Show that from one of these moves Sarah can ensure that she wins, but from the other Paul can ensure that he wins.

(ii) If the game is played from the beginning and Sarah has the first move, then she can always win the game if she plays correctly.
Explain how Sarah should play in order to be sure of winning.

Figure 18: *Knowledge and abilities to be tested* from PPN

Knowledge and abilities to be tested

The following list is intended to provide a general indication of the knowledge and abilities which the examination will be designed to test.

1. Knowledge of mathematical notation, terminology, conventions and units. The language and notation of sets together with the ideas of a mapping and a function are basic to the syllabus.
2. The ability to understand information presented in verbal, graphical or tabular form, and to translate such information into mathematical form.
3. The ability to recognise the mathematical methods which are suitable for the solution of the problem under consideration.
4. The ability to apply mathematical methods and techniques.
5. The ability to manipulate mathematical expressions.
6. The ability to make logical deductions.
7. The ability to select and apply appropriate techniques to problems in unfamiliar or novel situations.
8. The ability to interpret mathematical results.

Strategically, the initiative was made possible by my membership of the Research Advisory Committee and the Mathematics Subject Committee of the JMB, through which a relationship was built that allowed innovation. I pointed out that, of their list of seven "knowledge and abilities to be tested" in mathematics (Figure 18), only two or three were actually assessed by the then-current types of examination task. I convinced the Board that it was worth improving on this. The year-by-year change approach was accepted, the Shell Centre found funds to develop the support materials and tasks for the "live examinations". The Board's chief examiner for mathematics was part of the development team.

The design and development methodology used is also of interest. The initial design approach was different in the two cases. PPN was designed by the Shell Centre team with a group of teachers who were active members of the Association of Teachers of Mathematics. ATM had, for many years, pioneered approaches to teaching non-routine problem solving and more open mathematical investigations. LFG was designed by Malcolm Swan, building on a decade of Shell Centre research and development work on "translation skills" (Burkhardt, 1981) by Claude Janvier, Alan Bell and Malcolm Swan (Janvier 1981, Bell & Janvier, 1981).

One tactical design feature is worth noting. Each of the units demanded significant changes from the normal teaching style of most teachers. Non-routine problem solving is destroyed if the teacher breaks the problem up into steps, or guides the student through the mathematics – yet these are standard teacher moves when students are having difficulty. Similarly, LFG is built around classroom discussion in which students explain and discuss each other's reasoning, not expecting answers from the teacher. Aware that many teachers would not read extensive notes, we decided that the essential style changes should be summarized as a few key points on one page – the inside-back-cover of the teacher's guide (Figure 19). Feedback from the trials indicated that this worked well. (The five-session professional development material, which took the teaching issues further, was probably not widely used in schools – though it was popular with mathematics advisers for use in professional development activities that they led.)

Figure 19: Teacher guidance notes from PPN (left) and LFG

Checklist for the Teacher

When pupils learn to solve problems, they have to learn how to decide *what* to do and *when* to do it. If someone always tells them what to do, they *won't* learn these skills for themselves.

Aim to provide less and less guidance as you get further into the course.

Use freely any hints that make children think about the way they are tackling the problem:

"What have you tried?"
"Well, what do you think?"
"What are you trying to do?"
"Why are we doing this?"
"What will we do when we get this result?"

Use sparingly, particularly later on, hints about which strategies they should use:

"What have you found out so far?"
"Have you seen anything that is like this in any way?"
"How can we organise this?"
"Let's draw up a table of results"
"Can you see any pattern?"
"Have you tried some simple cases?"
"What examples should we choose?"
"How can we start?"
"Have you checked if that works?"

Avoid any hint referring to the particular problem:

"Do you recognise square numbers?"
"Explore it like this."
"Why don't you try using 3 counters?"

Classroom Discussion Checklist

These suggestions have been found useful in promoting useful and lively discussions in which pupils feel able to exchange ideas and examine hypotheses. Pupils are usually only *able* to participate if they have done some preliminary talking about the issues in pairs or small groups. Remember to allow time for this.

During classroom discussions, a teacher's role should:-

Mainly be that of a "Chairperson" or "Facilitator" who:-

— Directs the flow of the discussion and gives everyone a chance to participate
— Does not interrupt or allow others to interrupt the speaker
— Values everyone's opinion and does not push his or her own point of view
— Helps pupils to clarify their own ideas in their own terms

"Listen to what Jane is saying"
"Thanks Paul, now what do you think Susan?"
"How do you react to that Andrew?"
"Are there any other ideas?"
"Could you repeat that please Joanne?"

Occasionally be that of a "Questioner" or "Provoker" who:-

— Introduces a new idea when the discussion is flagging
— Follows up a point of view
— Plays devil's advocate
— Focuses in on an important concept
— Avoids asking 'multiple', 'leading', 'rhetorical' or 'closed' questions, which only require monosyllabic answers

"What would happen if . . ?"
"What can you say about the point where the graph crosses the axis?"

Never be that of a "Judge" or "Evaluator" who:-

— Assesses every response with a 'yes', 'good' or 'interesting' etc. (This often prevents others from contributing alternative ideas, and encourages externally acceptable performances rather than exploratory dialogue.)
— Sums up prematurely.

"That's not quite what I had in mind".
"You're nearly there".
"Yes, that's right".
"No, you should have said . . ."
"Can anyone see what is wrong with Jane's answer?"

The development process was also unusual. The first round of trials [36] was based on detailed classroom observation by a team of observers of about six teachers teaching the whole unit. The feedback meetings were based on detailed reports:

- First, on the style of each teacher, without and with the new materials, as seen by the observer concerned and any cross-observers; then
- Working step-by-step through the material, all the observers described what happened in 'their' classrooms.

To limit the amount of discussion that so easily runs on when consensus on design details is sought, I developed the principle of *design control* – while empirical feedback and design suggestions are strongly encouraged by the session chair, there is no search for consensus; feedback is absorbed and decisions taken by the lead designer of the units, in this case Malcolm Swan.

This approach to developmental feedback is, of course, much more expensive than, for example, relying on samples of student work alone. The modules were an example of "slow design" (de Lange, 2008). Each took about a year to develop and cost in all around $20,000 per lesson.

These materials had significant impact. The modules were bought by most of the schools that used this examination. The student responses to the tasks in the actual examination showed a reasonable range of performance. Since this was a new area of performance, it is no surprise that the level was much higher than in the exploratory tests at the beginning of the project.

This kind of 'switch on' gain is educationally both valid and valuable – the students acquired important new skills and the board's examination reflected more of their stated goals. There is a lesson in strategic design here. In contrast to attempts to raise standards in familiar areas of performance (adding fractions, using percentages, etc.), the introduction of important *new* areas, previously missing in examinations, if it is done well almost guarantees substantial success.

Again, this innovation foundered because of unconnected events – after only two years, the assessment at age 16 was restructured under larger organizations. Characteristically, such administrative changes absorb all the attention and energy of the bodies concerned for several years – working out the new arrangements and, incidentally, suppressing other innovations. It did not prove possible to carry over the relationship with the JMB to its broader successor body (NEAB, now AQA).

When policy makers weigh the likely benefits of reorganization, their speciality, they rarely recognize the cost – not only in disruption but in stopping ongoing improvement.

Some of the task types we introduced have persisted in other examinations, though usually in a more routine form. The "replacement unit" approach to step-by-step improvement has been successfully used by other designers – though often without change in high-stakes examinations, limiting the impact.

Comments on these successes

These examples show certain common features that may be more general:

- While a funding agency provided the opportunity, the design initiative came from a group with many years experience and recognized skill in the design of innovation, a deep understanding of the area they chose, and knowledge of the results of its long-term research agenda.
- The team was aware of the strategic design challenge and addressed the needs and concerns of all the key constituencies centrally involved – successfully, at least for a while.

- Well-aligned changes in the system's high-stakes assessment, when achievable, proved a powerful lever for increasing impact.
- Both an immediate "backlash" and a natural process of gradual erosion are to be expected; reflecting resistance to change, they need to be covered by persuasive arguments as well as evidence, and by ongoing "engines for improvement".

Any innovation in an education system is exposed to political and other events that will "blow it off course" (in a way that, for example, medicine and engineering are not). This is likely to remain true unless and until politicians and the public are persuaded that systematic research, design and development produces better solutions than the "common sense" that so often determines policy.

5. Principles for strategic design

In this section, I move on from examples and analysis to suggest some principles for strategic design. While none is essential for substantial, beneficial, ongoing impact on the system, they each seem to make it more likely.

Any such principles must, at this stage, be tentative; my hope is that they will be useful as a focus for the discussion that the paper aims to stimulate. To that end, I complement the principles with a list of issues – questions that need investigation as we try to understand better, and to improve, the interaction of innovative designs and the education systems they seek to improve. Finally I summarize the top-level goals that this work implies and some immediate actions that would forward them.

Strategic design principles

The following seem to be features of successful examples, while they have been neglected in the design of other innovations that failed. (They are phrased in the imperative.)

System awareness: Seek to understand the dynamics of the system you seek to improve, in all its interacting parts, and use it to guide the strategic design of the innovation.

Realism: Study the system as it is, not as it is intended to be, and the forces that shape decisions and actions of all the key groups, from politicians, parents and the media to teachers and their students; don't assume resources that have not been available without valid assurances that they will be.

Targeting: Be clear and specific about improvement aims, and the groups of users you are designing for – development should reconcile the goals and outcomes for those groups.

Alignment: Try to ensure that the set of tools and processes you develop form a coherent whole, in themselves and in interaction with the rest of the system – all the key players should be aware and "on board".

Robustness and flexibility: Since unexpected shocks to your plans are inevitable, try to design the set of tools and processes so that various elements can function independently in a range of contexts of use. For example, design so as to avoid "lethal mutations" (Brown & Campione, 1996) and to create designs that "degrade gracefully" (Walker, 2006)

Consensus building: Seek consensus on goals and entailments prior to design and throughout the development process – a profession that speaks with one voice has more influence on policy than one where diverse opinions reach policy makers. Consensus does not just happen; it often needs to be built through explicitly designed processes.

Communication and marketing: Be aware that any large-scale impact of your work will be influenced by the public, guided by the media. Improve your communication skills with these groups, and your network of contacts.

Space for excellence in tactical and technical design: Work to retain as much space as possible for the creative talents in your design team, and the systematic development that refines the products – good strategic design is worthless without them.

"We must educate our masters": Seek to make policy makers, funders, and designers aware of the crucial role strategic design will play in the success of the enterprise in turning its goals into large-scale impact.

Big challenges need big teams: The range of skills needed to carry through a design and development program, with high-quality in all its aspects, needs to be reflected in the design team – often, particularly for large scale developments, only a multidisciplinary team can understand and work with the various communities that will interact with the product.

Strategic design issues

At a more detailed level, there are various issues in strategic design which merit systematic investigation [37]. Referred to above, they all relate to aspects of choosing a *model of change* for an innovation. The appropriate choice will depend on features of the existing system, and on the resources likely to be available to support the change. A key variable, usually neglected, is *the pace of change in their practice* that the crucial performers (often teachers) are likely to be able to achieve, without corrupting the intentions of the innovation – a too-common outcome. Such issues include:

How big a step? How ambitious should a change be? [38] For a given level of support, if the step is too large, few will take it without stumbling; if it is too small, why bother? If we are ambitious, can we define and support a *pathway of progress* for the key performers, particularly teachers, so the latter can gradually move to match our ambitions in their classrooms?

'Big bang' v incremental change Should we seek to achieve our goals as a single major change (e.g. introducing a new curriculum, as in 4A and 4B, or incrementally, as a planned sequences of changes (as in 4D)?

Small steps can be less expensive, more easily sold, and more digestible to users. A comprehensive reform is more conventional, and more satisfying to many, including politicians who like to "solve problems" (though other fields, like medicine and engineering, move incrementally). The trade-offs are fairly clear; the best choice less so – and system dependent.

Time scales How can we meet short term political thinking and achieve anything useful? We have noted the fundamental mismatch between the time-scale of politics and that of significant educational improvement. Politicians need to show results well before the next election. Education systems are built around professionals, skilled in aspects of their work through well-grooved practices; changes in those practices, particularly those requiring new skills, take time – for the system as a whole, typically a decade or more. For example, the recognition in the US in the early 1980s of the need for change in mathematics education led, through the NCTM Standards in 1989, to the NSF-funded curriculum projects, whose products began to impinge on the textbook market around 2000 – the process of institutionalization, in which these curricula become the accepted norm, still continues [39].

I believe there is opportunity for creative thinking on ways to reconcile these two timescales. One approach is planned incremental change, which can provide politicians with a sequence of year-by-year successes to claim within a decade-long improvement schedule that learns as it proceeds. I expect that there are others.

Standard slots v new opportunities Should we try to improve an existing entity (e.g., a school subject or an examination) or to get a new one accepted, as a replacement or as an alternative? This dilemma has faced many innovators. For example, statistics education, which has developed internationally as a problem-based subject built on interpreting real data, has long been unhappy with being seen as part of school mathematics (though that needs much more of the same approach!). This discomfort remains but attempts to get a separate subject slot in the timetable have had limited success. (It is available as an option in the UK for age 14 upwards.)

That otherwise-excellent book *Mathematics and Democracy* (Steen, 2001) even suggests that *quantitative literacy* [40] should be taught separately from Mathematics – without questioning why society should give so much curriculum time to a secondary school mathematics curriculum that is both non-functional and non-motivating for most students, and the adults they become.

Cross-subject teaching of subjects, though often advocated, has proven even harder to establish; schools are still organized around subject slots, each with its own agenda.

The dilemma, and the trade-offs, are clear. Little or no impact in a pure form versus wider impact of a debased version. Some novel approaches have been tried, with success on a small scale – for example, a whole-school project day every week or two. It is worth looking for others.

All these issues of strategic design deserve more study and experiment.

6. Improving strategic design

Finally, I propose a set of long and medium term goals, together with immediate actions that seem likely to forward their achievement.

Long term goals

Recognition by policy makers that education can and should become a research-based field, like medicine where:

- Insight-focused analytic research on working systems is the best route to diagnosis of problems and their likely causes;
- A long term agenda for improvement is complemented, as in other fields, by a regularly-reviewed sequence of steps along the way with sensible timescales;
- Good engineering, integrating insights from prior research and development, design research, systematic development and evaluation in depth, will produce the most effective solutions;
- Strategic design of their initiatives should be as professional as the tactical and technical design already (sometimes) are, using the same methodology; and
- Much better evaluation of products and initiatives in education[41], covering in some depth both the various outcomes and the conditions under which they were achieved.

For this we will need:

- More researchers choosing projects and using methodologies that provide the in-depth evaluative evidence that policy makers need on products and processes that are widely available, yielding reliable evidence on “what works, how well, under what circumstances”[42]; and
- More people trained in engineering research methods to design and develop robust solutions.

In mathematics education we will need to learn to emulate science education in developing:

- Effective machinery for building a consensus on what is needed, and the steps along that road, leading to;
- Unified recommendations for innovation that reflect government realities.

Medium term goals

Recognition by policy makers that (as in health care, for example):

- High-stakes targets (i.e. tests) for (teachers and schools) can distort priorities, ensuring that the implemented curriculum in most classrooms is no better than what is tested. The good news (as in Section 4) is the substantial evidence that better tests can be an effective lever for improvement.

- What is achievable within the timescale required and resources available is:
    - An empirical question that can only be reliably answered by imaginative design, systematic development and evaluation in some depth;
    - Usually much less than is desirable – or is promised by "experts" who are keen to please government but have no valid evidence for what they recommend; and
    - Will require funding with at least a few–year timescale, involving competitive design groups and independent evaluation with agreed criteria and methodologies (c.f. NICE in health care).
    - After so many failures from "obviously needed" reforms, there is political capital to be gained from a sensible research-based approach.

Short term actions

Over the next year or two, we can move to strengthen the case for the above goals by:

- Identifying, as in section 4, examples of successful design, then studying the various aspects of their strategic design in some depth, and in comparison with parallel innovations where these are available;
- Identifying, and specifying in some detail, alternative models of change, analyzing their key features and the expected cost-benefit analysis;
- Refining and strengthening evidence of payoff from giving medium-term support to high-quality design teams with proven track records in well-defined areas; and
- Developing effective channels for communication and influence on policy makers – one of ISDDE's prime goals to which this journal is a contribution.

There is much to be done to review, strengthen and implement these proposals. Better understanding of strategic design, and how it interacts with other aspects will surely be part of it. I believe such an enterprise is worth increased attention and, insofar as it succeeds, will forward both learning and teaching and the development of the profession of educational design.

It is worth remembering von Clausewitz' definition of strategy as the ability to "make the best use of the few means at our disposal".

Acknowledgements

The thinking behind this paper has built on discussions over many years with fellow designers, project leaders and funders. I am particularly grateful to Phil Daro, Alan Schoenfeld, Mark St. John, Glenda Lappan, Janice Earle, Quentin Thompson and my colleagues in the Shell Centre, Daniel Pead and Malcolm Swan. Others, notably Susan McKenney, have provided helpful suggestions. The limitations are my own.

I hope that responses to this paper will help us collectively to move forward in our understanding of strategic design, how to improve it, and how to persuade funders that it is an area of design that needs both imagination and systematic development.

References

Barnes, M., Clarke, D., and Stephens, M. (2000). Assessment: The engine of systemic curricular reform? *Journal of Curriculum Studies,* 32(5), 623-650.

Bell, A., Janvier, C. (1981) The Interpretation of Graphs Representing Situations. *For the Learning of Mathematics* 2(1) pp34-42.

Black, P.J., & Atkin, J. (Eds.), (1996). *Changing the subject: Innovations in science, mathematics and technology education*. London: Routledge.

Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education* **5**(1) pp. 7-71.

Black, P. & Wiliam, D. (2001). *Inside the black box*. London: Kings College

Black, P. (2008). Strategic Decisions: Ambitions, Feasibility and Context. *Educational Designer*, 1(1). http://www.educationaldesigner.org/ed/volume1/issue1/article1

Bowland (2008). *Bowland Maths - Key Stage 3* (Website and DVD-ROM). Bowland Charitable Trust, Blackburn, UK. http://www.bowlandmaths.org.uk/

Brown, A. L., & Campione, J. C. (1996). Psychological learning theory and the design of innovative environments: On procedures, principles and systems. In L. Shauble & R. Glaser (Eds.), *Contributions of instructional innovation to understanding learning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Burkhardt, H., (1981). *The Real World and Mathematics*, Blackie: Glasgow, revised Nottingham: Shell Centre Publications 2000.

Burkhardt, H. (1990). On Specifying a National Curriculum, in *Developments in School Mathematics Worldwide*, Wirszup, I & Streit, R. (Eds.) Chicago: University of Chicago School Mathematics Project.

Burkhardt, H. (2006). From design research to large-scale impact: Engineering research in education. In J. Van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research*. London: Routledge.

Burkhardt, H. (2008). Firmer Foundations for Policy Making: The unrealised potential of 'engineering research' in education. Paper prepared for Department of Children School and Families. http://www.mathshell.org/papers/pdf/dcsf_strategy.pdf

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving Educational Research: towards a more useful, more influential and better funded enterprise. *Educational Researcher, 32*, 3-14.

Clarke, D. J. & Stephens, W. M. (1996). The ripple effect: the instructional impact of the systemic introduction of performance assessment in mathematics. In M. Birenbaum and F. Dochy (eds), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 63-92). Dordrecht, The Netherlands: Kluwer.

Cockcroft (1982). *Mathematics Counts*. London: HMSO.

Cohen, D.K., Hill, H.C. (2001). *Learning Policy: When State Education Reform Works*. Yale University Press, New Haven, CT

Crowther (1959). Report 15-18 *Half our future*. A Report of the Central Advisory Council for Education, HMSO: London.

Curtis, D. & Denton, R. (2003). *The Authentic Performance-based Assessment of Problem-solving*. Station Arcade, South Australia: National Centre for Vocational Education Research.

de Lange, J. (2008). Invited talk at the ISDDE annual meeting. June 29-July 2. Egmond aan Zee, the Netherlands.

DfES. (1988). *Proposals of the Secretary of State for the National Curriculum in Mathematics*, Department for Education and Science, London: HMSO.

Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Goodlad, J. (1994). Curriculum as a field of study. In T. Husén, & T. Postlethwaite (Eds.), *The international encyclopedia of education* (pp. 1262-1267). Oxford: Pergamon Press.

Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks: Corwin Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: theory and practice, 8*(3/4), 381-391.

Janvier, C. (1981). Use of Situations in Mathematics Education. *Educational Studies in Mathematics* 12 113-122.

Joyce, B., & Showers, B. (1980). Improving science teaching: The messages of research. *Educational Leadership, 37*, 379-385.

Joyce, B., & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal* (2nd ed.). White Plains, NY: Longman.

Lappan, G., Phillips, E. (2009) A Designer Speaks. *Educational Designer*, 1(3). http://www.educationaldesigner.org/ed/volume1/issue3/article11

Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA:

Corwin Press.

MARS (2001-). Crust, R., Burkhardt H. and the MARS team. *Balanced Assessment in Mathematics*. Annual tests for Grades 3 through 10. 2001–2004. Monterey, CA: CTB/McGraw-Hill; 2005- Nottingham: Shell Centre Publications

McCrae, B. (1995). An evaluation of system-wide assessment of problem solving at Year 12 by report and related test. In *Proceedings of the 18th Annual Conference of MERGA*.

Michigan Mathematics Standards (2005). http://www.ixl.com/math/standards/michigan/ Retrieved 2 November, 2009.

NCTM (1989). *Curriculum and Evaluation Standards for Mathematics*, National Council of Teachers of Mathematics, Reston VA: NCTM

NGA, CCSSO (2009). *Common Core Standards Initiative*, National Governors Association/Council of Chief State School Officers, Washington DC. http://www.corestandards.org/ Retrieved November 2009

Ofsted (2006) Evaluating Mathematics provision for 14-19 year-olds. Oftsted, London. http://www.ofsted.gov.uk/Ofsted-home/Publications-and-research/Browse-all-by/Education/Curriculum/Mathematics

Ofsted (2008) Mathematics: understanding the score. Ofsted, London. http://www.ofsted.gov.uk/Ofsted-home/Publications-and-research/Browse-all-by/Education/Curriculum/Mathematics

Phillips, R.J., Burkhardt, H., Fraser, R., Coupland, J., Pimm, D., & Ridgway, J. (1988). Learning activities & classroom roles with and without the microcomputer. *Journal of Mathematical Behavior, 6*, 305–338.

Plomp, T. (1982). *Onderwijs technologie: Enige verkenningen* [Educational technology: Some explorations]. Inaugural address. Enschede: University of Twente.

QCDA (2007). *Mathematics Key Stage 3 Programme of Study*. Qualifications and Curriculum Development Agency, London. http://curriculum.qcda.gov.uk/key-stages-3-and-4/subjects/mathematics/keystage3/ Retrieved November 2009.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

Schoenfeld, A. H. (2002). Research methods in (mathematics) education. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 435–488). Mahwah, NJ: Erlbaum.

Shell Centre (1984). Swan, M., Pitt, J., Fraser, R. E., & Burkhardt, H., with the Shell Centre team, *Problems with Patterns and Numbers*. Joint Matriculation Board, Manchester, U.K.; reprinted 2000, Shell Centre Publications, Nottingham, U.K. http://www.mathshell.com/scp/index.htm.

Stacey, K. (1995) The challenges of keeping open problem solving open in school mathematics. *Zentralblatt fur Didaktik der Mathematik* ***95*** 62 – 67.

Stacey, K. & Groves, S. (1985) *Strategies for Problem Solving*, Melbourne: Latitude Publications

Stephens, M., & McCrae, B. (1995). Assessing problem solving in a school system: principles to practice. *Australian Senior Mathematics Journal, 9*(1), 11–28.

Steen, L.A. (ed) (2001) *Mathematics and Democracy – The Case for Quantitative Literacy*. National Council on Education and the Disciplines, USA. http://www.maa.org/ql/mathanddemocracy.html

Swan, M. with the Shell Centre team: (1986). *The Language of Functions and Graphs*, Manchester, U.K.: Joint Matriculation Board, reprinted 2000, Nottingham, U.K.: Shell Centre Publications. http://www.mathshell.com/scp/index.htm.

Van den Akker, J., Gravemeijer, K., McKenney, S. & Nieveen, N. (Eds.) (2006). *Educational design research.* London: Routledge.

Verhagen, P.W. (2000). *Over het opleiden van onderwijskundig ontwerpers.* [On educating educational designers] Inaugural address. Enschede: University of Twente.

Victorian Board of Studies (1995). *VCE Mathematics: Specialist – Official Sample CATs*. Blackburn: HarperSchools**.**

Walker, D. (2006). Toward productive design studies. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.) *Educational design research.* London: Routledge.

Footnotes

[1] The contrast with medicine, for example, is stark. One cannot imagine penicillin getting lost.

[2]Apart from their military connotations, these terms parallel those that Alan Schoenfeld coined in his analysis of mathematical problem solving (Schoenfeld, 1985). This is appropriate because design *is* a type of problem solving. Schoenfeld added a fourth metacognitive aspect, *control* – the monitoring and guiding of the problem solving process – which is reflected in *design control*..

[3] I have mainly contributed to the other aspects.

[4] I will say little about this important area because, particularly in mathematics education, the choice of learning goals is often confused with the pedagogical

question of how that learning should be achieved. The latter is the focus of most of the controversy.

[5] The description of the camel as “a horse designed by a committee”, while a slander on that admirable beast, captures this important point.

[6] “Goodhart’s Law” states that “when a measure becomes a target, it ceases to be a good measure” – essentially because targets promotes gaming and other distortions described here. Dylan Wiliam’s version is “The higher the stakes, the worse the assessment”. There is evidence here that this, while commonly true, is not inevitable.

[7] Ironically, what usually happens in practice is the reverse of parents choosing schools for the kids; because of limits of capacity in each school, popular schools choose their students.

[8] In health care it is now well-recognized that unbalanced targets distort clinical priorities. For example, an earlier emphasis in the UK on reducing maximum waiting times led to the treatment of some more urgent cases being delayed.

[9] If an English Language test were relabeled Mathematics, its psychometric “reliability” would be unchanged. The statistical tools used measure consistency and levels of difficulty; *they say nothing about what is being assessed*.

[10] The Cockcroft Report (1982) defined the purpose of good assessment in these terms.

[11] I have avoided the standard terms, *validity* and *reliability*, because their usual non-technical meanings are distorted to allow statistical definitions that amount simply to *consistency* – between individual items and the test and between supposedly equivalent tests.

[12] When the National Curriculum was being developed, a senior UK policy maker said to me: “Well, with maths, it’s things you can either do or you can’t, isn’t it?” and went on to impose this checklist approach on the Working Group. For English Language, which politicians and policy makers understand much better, essays and other extended writing, not just vocabulary lists and grammar rules, are central in both the standards and the tests. There are no substantial tasks in the Mathematics standards or tests.

[13] Though more sophisticated concepts, such as the formalism for arithmetic progressions, can be used profitably in this task, most of the interesting results can be found without them – and few 16-year–olds can use them in non-routine problems, even when they can in routine exercises.

[14] For example, examination tasks in Euclidean Geometry consisted of two parts: a “proof” of a theorem that the student was expected to have learnt, followed by a “rider” that involved using the theorem, among other things, to solve a non-routine extension.

[15] This approach implicitly recognizes the weakness of models of performance based entirely on analytic descriptions of the elements of the domain. However, many in authority, inside and outside the field, find it surprisingly difficult to accept specifications that are partly based on exemplars.

[16] 40 pages of varied task exemplars (typically 5 to 20 minutes) were included in the original version of the National Curriculum (DfES 1988), designed by the Government's Mathematics Working Group and circulated for comment. The removal of the exemplar tasks from subsequent revisions was never explained; it was probably their lack of one-to-one alignment with the detailed level criteria. The length of each of the test items that emerged is about 90 seconds.

[17] A comforting phrase that is often used in England.

[18] For example, some schools exclude students from the high-stakes Grade 12 A-level examination in those subjects where they did not do well in the Grade 11 AS examination, even though that subject may be important to a student's future plans.

[19] Jan de Lange, at the 2008 ISDDE Egmond conference, used the phrase "slow design" as the route to excellence.

[20] As in architecture, where competitions for important buildings are common.

[21] The UK Government had a 'grid' system across the various departments of state designed to ensure that there was a new announcement every few days to keep the media happy.

[22] In an example from health care of political decisions under outside pressure, the USA and New Zealand allow pharmaceutical companies to market their drugs directly to consumers. Patients, despite their lack of diagnostic expertise, are increasingly demanding certain treatments.

[23] "The Green Book", http://www.hm-treasury.gov.uk/data_greenbook_index.htm, lays down procedures to be followed by all parts of central government.

[24] I am not trying to suggest that all medical procedures are research-based but the ongoing movement in that direction is a good model for education. (In both fields, practitioners also respond to societal demands, however unfounded – e.g. antibiotics for virus infections, long division by hand!)

[25] I am grateful to the lead designers for their help the with facts; the analysis is mine.

[26] The Nuffield Foundation has a remarkable record of successful innovation in science education. The Foundation has always paid attention to the system issues involved.

[27] At that time students' A-level certificates bore the signature of the Secretary of

State!

[28] They were awarded the ISSDE Prize for Educational Design in 2008 for *Connected Mathematics*.

[29] Broader spectrum tests are available (MARS, 2000–) and are used in some school districts, mostly in California, in addition to the state tests. Their influence has been mostly indirect.

[30] This is one reason behind the comment that US curricula tend to be "mile wide, inch deep."

[31] The *customer*, usually the school district leadership, makes the decision to buy materials; the *clients*, teachers and students, use them.

[32] In the absence of the substantial empirical effort that, as in medicine, would be needed to collect reliable evidence, this was the usual "evaluation by inspection".

[33] Mathematics was broken into four "subjects"; students would choose at most two a year.

[34] Their work had been recognized in their appointment by the International Program Committee of the 5th Internal Congress on Mathematical Education as Australian Organizers of the Problem Solving theme at the 1984 Adelaide Congress (Alan Schoenfeld and I shared this with them).

[35] Malcolm Swan, the lead designer, was awarded the ISSDE Prize for Educational Design in 2008 for this module. "The Red Box" is still used and talked about across the English-speaking world.

[36] The second round of development was conventional. A representative sample of about 30 classrooms trialled the materials, with teachers reporting back on their experience and sending sample student work.

[37] Some of these issues have been discussed in the Dutch literature, e.g by Plomp (1982), Verhagen (2000), van den Akker, Gravemeijer, McKenney and Nieveen (2006).

[38] The 13 projects in mathematics supported by NSF, of which 4B is one, made a wide range of choices – from modest change to near-revolution.

[39] This 25-year timescale, from an initiation event to systemic change, is not atypical across fields, exemplified by penicillin and the vacuum cleaner among other revolutionary innovations.

[40] Variously called *mathematical literacy* by PISA and many others, *functional mathematics* in the UK, *quantitative reasoning*, and *numeracy* (in its original meaning, see Crowther Report 1959), so often now corrupted to mean basic

skills in arithmetic.

[41] In the UK the National Institute for Health and Clinical Excellence (NICE) evaluates drugs and medical procedures for effectiveness and, somewhat controversially, for cost-effectiveness. Many countries, faced with rapidly rising cost of new treatments, are considering such systems.

[42] This emphasis is very different from current fashions in research "quality" (Burkhardt and Schoenfeld 2003). More basic research should be seen as long term, also demanding evidence of the generalizability (Schoenfeld 2001) of results – in particular, how far they are valid and robust in the domain of the intended design application.

About the Author

Hugh Burkhardt has been at the Shell Centre for Mathematical Education at the University of Nottingham since 1976, as Director until 1992. Since then he has led a series of international projects in the UK and the US, including Balanced Assessment, the Mathematical Assessment Resource Service (MARS), and its development of a Toolkit for Change Agents.

He takes an 'engineering' view of educational research and development - that it is about systematic design and development to make a complex system work better, with theory as a guide and empirical evidence the ultimate arbiter. His core interest is in the dynamics of curriculum change. He sees assessment as one important 'tool for change' among the many that are needed to help achieve some resemblance between goals of policy and outcomes in practice. His other interests include making mathematics more functional for everyone through teaching real problem solving and mathematical modeling, computer-aided math education, software interface design and human-computer interaction. He graduated from Oxford University and the University of Birmingham where, alongside research in theoretical physics and undergraduate teaching, he first developed his work on teaching the uses of mathematics to help solve everyday life problems. He remains occasionally active in elementary particle physics. He founded and is Executive Chair of the International Society for Design and Development in Education.

Burkhardt, H. (2009) On Strategic Design. *Educational Designer,* 1(3).
Retrieved from: http://www.educationaldesigner.org/ed/volume1/issue3/article9