

Design and Development for Large-Scale Improvement

Hugh Burkhardt and Malcolm Swan

Abstract This chapter describes the Shell Centre team’s “engineering research” approach to the improvement of practice through researched-based design and development of tools for teaching and learning mathematics, for professional development and for supporting large-scale change. The contributions of projects over the past 35 years to the development of design principles and tactics are outlined and illustrated. The roles of tasks of different kinds in learning and assessment are explained, with particular reference to the design of tests, and of formative assessment lessons for concept-development and problem solving. The chapter concludes with a look at the barriers to turning success at classroom level into large-scale change—and how this challenge can be tackled.

Keywords Design · Engineering · Strategy · Formative · Assessment · System

Introduction

The creation of the Emma Castelnuovo Award by ICMI is an important milestone in linking research and practice in mathematics education. The core focus of academic research is on deeper understanding of a field and its phenomena: in mathematics education, exceptional contributions are recognized by the Felix Klein and Hans Freudenthal Awards. But other fields with direct impact on people’s lives—medicine and engineering, for example—balance this insight-focused research with research-based design and development of new products and processes that enable practitioners to tackle more effectively the problems of practice. A large part of medical research, for example, is focused on developing new and more effective

Emma Castelnuovo Award.

H. Burkhardt (✉) · M. Swan (Deceased)
School of Education, Shell Centre, University of Nottingham, Jubilee Campus,
NG8 1BB, Nottingham, UK
e-mail: Hugh.Burkhardt@nottingham.ac.uk

medicines, devices and procedures. Equally, our lives are filled with products of engineering research and development that embody the new fundamental insights research provides. No such balance exists in education, where impact-focused research remains relatively rare. Its importance is recognized by this new award.

This chapter is primarily an account of what we and our colleagues at the Shell Centre have done over the last 35 years to develop and exemplify this “engineering research” approach (Burkhardt, 2006). Towards the end, we will discuss strategic changes in mathematics education research that would encourage a better balance of *insight-focused* and *impact-focused research*, giving the direct serving of practice the priority it deserves.

The Shell Centre Approach

The Shell Centre for Mathematical Education was founded as a professional development centre in 1968 by Nottingham’s professors of pure and applied mathematics, Heini Halberstam and George Hall. The vision at that time was that improving teachers’ understanding of mathematics and its applications was the key to improving student learning. By the time one of us (HB) was appointed director in 1976, it was becoming recognized that the challenges were much broader than ‘knowing more maths’, so a radically different ‘brief’ for the Shell Centre was agreed:

To work to improve the teaching and learning of mathematics regionally, nationally and internationally.

This ambitious challenge had a chain of implications:

- The focus should be direct impact on practice in classrooms.
- Large-scale impact can only be achieved through reproducible materials.
- Developing these well needs engineering-style research, which other fields have shown can produce both better products and new research insights.
- Good engineering implies a focus on design—strategic, structural, technical—and on systematic development in appropriate contexts.

This led to a search for outstanding designers: the other author (MS) was invited to join the Centre in 1979.

What distinguishes these different aspects of design?

Strategic design (Burkhardt, 2009) is concerned with the “fit” of a design with the system it aims to serve: finding “points of leverage”, for example high-stakes testing; devising models of change that work well; guiding policy in a way that satisfies the needs of all the key groups, including policy makers. Poor strategic design is a common source of failure of initiatives to achieve their goals.

Structural design aims to ensure that a tool fits both the ‘user’ and the ‘job’ being addressed—just as a knife has a handle and a cutting edge, so materials for teaching problem solving should support the teacher-user in helping the students to develop strategies for solving non-routine problems.

Technical design of the product relies on a combination of input from prior research and design creativity that injects the “surprise and delight” that, along with a sound research basis, epitomizes excellence and gives pleasure to users, both teachers and students.

The other essential element for turning designs into products that are both educationally ambitious and reliably effective in the hands of diverse users is the same as in any impact-focused research-based field:

Systematic development through trials in realistic conditions with the rich and detailed feedback needed to guide improvement. For us this has meant direct observation of trials with reports to the designers based on protocols structured to focus on the key events.

These principles have guided a sequence of linked design and development projects through which the Shell Centre team has developed tools and processes for classroom teaching and learning, formative and summative assessment, and teacher professional development. In the next section we explain the key roles that tasks play. In Section “[Developing Design](#)” we describe how specific projects have led us to identify new design principles and tactics. Sections “[Developing Conceptual Understanding and Logical Reasoning](#)” and “[Developing Strategies for Problem Solving](#)” describe the design of formative assessment lessons to support concept development and problem solving, respectively. In the last decade we have come to see that major obstacles to progress lie at levels “above” the classroom. Section “[Tools for Supporting Systemic Change](#)” describes our work on tools and processes to advance systemic improvement.

Descriptions alone cannot adequately communicate design ideas or products; exemplification is essential—but, in a book like this, inevitably brief. The website emma.mathshell.org gives examples in full, section by section, along with sketches of all the main Shell Centre Projects.

Building an International Community

The above approach is broadly shared by some other design teams around the world, though it remains rare in the huge body of education research. While there have always been international exchanges of ideas, it seemed to us that the profession would benefit from coming together more formally to share common challenges and opportunities. After discussions over a decade or so, we launched the International Society for Design and Development in Education (ISDDE) at a conference in Oxford in 2005. Since then annual conferences have been held in different parts of the world. The Society currently has about 100 Fellows. Its goals are to:

- Build a design community—this now exists
- Raise standards in design and development—there has been progress, learning together
- Increase influence on policy—this remains a, perhaps *the*, major need and challenge.

Educational Designer was established by ISDDE to share expertise. We decided that a peer-reviewed e-journal format was best because it allows articles to combine relatively brief and readable text with the rich exemplification needed in talking about design, accessed through internal links. Much of the design detail that is perforce squeezed out of this chapter can be found in articles in the journal.

ISDDE is focused on education in mathematics, science, engineering and technology. There are fundamental reasons why design is more important here. In teaching the humanities, teachers master a modest number of lesson *genres* into which they insert texts which they choose from the varied literature of their subjects, producing an infinite variety of lessons. The original literature in subjects such as mathematics and science is too technical for use in school—hence the need for the detailed design of coherent, linked lessons that bring students to grips with various aspects of understanding and doing mathematics—not just *lesson genres*, though these are important.

Tasks in Mathematics Education

Tasks play at least four important roles in mathematics education:

- **Providing ‘microworlds’ for investigation:** as a stimulus for learning; for developing understanding; for learning strategic methods for tackling complex non-routine problems.
- **Summarizing curriculum goals.** Analytic domain descriptions—“national curricula” or “standards”—are highly ambiguous; complementing them with an exemplar set of tasks covering the target types of performance makes the learning goals much clearer.
- **Assessing students’ performance** through tests and coursework for monitoring progress, for selection or for accountability purposes, or through formative assessment in classrooms.
- **Providing targets for performance.** Mathematicians set research targets in terms of tasks: Prove Fermat’s last theorem or the 4-colour map problem, solve ‘Hilbert problems’ or the ‘Travelling salesman problem’. Teachers use tasks from ‘past exam papers’, often over-concentrating their teaching on those task-types.

Tasks and their design is a recurring theme throughout our work.

Task Difficulty

It is important to choose tasks that the students find interesting and challenging—but not impossible! It is known from research that the difficulty of a task depends on various factors, notably its:

- **complexity**—the number of variables, the variety and amount of data, and the number of modes in which information is presented, are some of the aspects of complexity that affect the difficulty a task presents.
- **unfamiliarity**—a non-routine task is more difficult than a similar task one has practised solving; the student has to understand the structure of the situation, work out how to tackle it, and do so while monitoring progress.
- **technical demand**—a task that requires sophisticated mathematics for its solution is more difficult than one that can be solved with more elementary mathematics.
- **student autonomy**—guidance from an expert (usually the teacher), or from the task itself (e.g., by structuring or “scaffolding” it into successive parts) makes a task easier than if it is presented without such guidance.

Assessments of student performance need to take these factors into account. For example, they imply that, in order to design a task for a given level of difficulty, a relatively *complex non-routine* task that students are expected to solve without guidance needs to be *technically* easier than a short exercise that develops or tests a well-practised routine skill. Problem solving tasks need to be *conceptually* easier than those that focus on mathematical concepts.

Rich tasks allow students at different levels to provide different correct responses. For such tasks, difficulty also reflects the level at which the student engages with the task. This is similar to the situation in English or History; the same essay question might be posed to a young student or to a college graduate, expecting quite different “good” responses.

‘Expert’, ‘Apprentice’ and ‘Novice’ Tasks

We have found it useful to distinguish three broad types of task: ‘expert’, ‘apprentice’ and ‘novice’ tasks. Each type has a different balance of sources of difficulty.

‘Expert Tasks’ are problems in the form they naturally arise, in the real world or within mathematics. Relatively complex and non-routine, if the students are to be able to solve them autonomously they must not be technically demanding. Figure 1 shows two expert tasks, accessible in some middle school classrooms—and also good with older students. The difficulty here comes mainly from the *complexity*, with various factors, not all stated, and *unfamiliarity* so the students have to work out what to do, then do it by constructing a chain of reasoning. ‘Table Tiles’


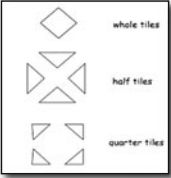

Table Tiles	Traffic Jam
<div style="display: flex; justify-content: space-around;">   </div> <p>Maria makes square tables, then sticks tiles to the top.</p> <p>Maria uses whole tiles in the middle, quarter tiles at the corners and half tiles on the edges.</p> <p style="text-align: right;">Describe a quick method for calculating the number of tiles of each type that are needed for any size of table top.</p>	<p>A queue on a two-lane freeway is 12 miles long.</p> <p>How many cars are in the traffic jam?</p> <p>Drivers have a two-second reaction time.</p> <p>When the accident clears, how long is it before the last car moves?</p> 

Fig. 1 Two ‘expert tasks’

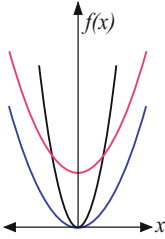
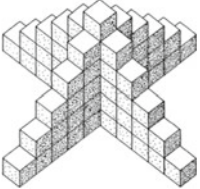
<p>(a) Algebraic Functions</p> <ul style="list-style-type: none"> • Factorise $x^2 + 3x - 4$ • Solve $3x + 5 = 21 - 5x$ • Write $\sin(A + B)$ in terms of \sin and \cos of A and B. • The three graphs show the functions, $y = x^2$, $y = x + k$, $y = kx^2$ where: $k > 1$ Label the graphs. 	<p>(b) A Skeleton Tower</p>  <p>How many cubes do you need to make a tower:</p> <ul style="list-style-type: none"> • 6 cubes high? • 20 cubes high? • n cubes high? <p>Explain your reasoning. Can you find another method?</p>
--	---

Fig. 2 a Novice tasks. b Apprentice task

involves detecting and describing patterns; it includes a “ramp of difficulty”: 4 quarter tiles provide the corners of any such table, while the number of half tiles is linear and of whole tiles is quadratic in the table size—reflecting the deep insight that corners are points, sides are lines and the centre is an area. ‘Traffic Jam’ is about proportional reasoning—arguably the most important modelling tool that students should learn to use in school.

‘Novice Tasks’ are short ‘items’ with mainly technical demand (as in Fig. 2a). Each is focused on a specific concept or skill, so they can be “up to grade”,

including content that has recently been taught. Novice tasks are designed to test recall of learned procedures. (Novices are learning the tools of the trade.) ‘*Apprentice Tasks*’ (e.g. Fig. 2b) are expert tasks with scaffolding added to guide the student in a series of steps, reducing complexity, non-routine-ness and student autonomy. (Apprentices learn to solve problems with expert guidance).

The difficulty in ‘Skeleton Tower’ lies mainly in deciding how to tackle the problem; this is scaffolded by the two specific examples. 6 cubes high is pictured; it can be done by counting which, if recorded, reveals a pattern that can then be extended numerically to 20. Without these steps, this would be an expert task requiring a verbal rule or formula—the last part. Note that expertise involves learning problem solving strategies (Polya, 1945; Schoenfeld, 1985; Swan et al., 1984), including “try some special cases” and “look for patterns and structure”—removing the need for scaffolding like this.

The difficulty of a task has ultimately to be determined by trialing the task with the target group of students. All assessment tasks, whether for use in the classroom or in summative tests, should be developed in this way. To summarize the key point, there is a “few year gap” between the mathematical concepts and skills that a student can use in short imitative novice tasks and those they can use autonomously in solving expert tasks. Students’ mathematical expertise is what matters beyond school yet, currently, the curriculum in many countries has only novice tasks—leading to a novice-level mathematics education. To develop factual knowledge, conceptual understanding and strategic competence, a world-class mathematics education needs substantial experience of all three kinds of task—novice, apprentice and expert—in both curriculum and assessment.

Learning Goals and Task Genres

We have learned a lot from ‘own language’ teaching. This seeks to develop technical fluency (spelling, grammar, syntax), to analyze and create texts in different genres (reports, letters, stories, poems, etc.) and to relate texts to social, historical and cultural contexts. Progress consists in being able to handle more challenging texts in more sophisticated ways. If you change “texts” to “tasks”, the top-level goals for mathematics are much the same.

We have recently come to develop a framework that looks in more detail at tasks in terms of the primary purposes of the learning activity they support, the genres of student activity that the task demands, and the type of product that results from the student reasoning involved. It is summarized in Table 1.

In the UK, the US and many other countries the first row, *facts and procedure development*, is dominant in assessment and in many classrooms. (The policy rhetoric is often broader). Facts and procedure are easy to assess through tests using novice tasks. Teaching and assessment for the other purposes, however, needs apprentice and expert tasks. *Conceptual understanding* requires chains of reasoning, connections and explanation—as do *problem solving and strategic competence*.

Table 1 An activity-genres framework for design

Purpose of the lesson	Process genres on tasks—the student:	Student products
Factual knowledge and procedural fluency (e.g. Fig. 2a)	Memorizes and rehearses facts, procedures and notations	Performance
Conceptual understanding and logical reasoning (e.g. Fig. 2b)	Observes and describes phenomena	Description
	Classifies and defines objects	Classification
	Represents, and translates between representations	Representation
	Justifies conjectures, procedures, connections	Explanation
	Identifies and studies structure	Analysis
Problem solving and strategic competence (e.g. Fig. 1)	Formulates models of situations	Model
	Solves non-routine problems	Problem solution
	Interprets and evaluates strategies	Critique

The design of such tasks requires a much broader range of partly-creative design and development skills than official test providers typically possess. Such tasks are therefore lacking in most tests—and therefore in most classrooms. The work described below exemplifies how this need for range and balance can be met. For example, in the Mathematics Assessment Project, described in Sections “[Developing Conceptual Understanding and Logical Reasoning](#)” and “[Developing Strategies for Problem Solving](#)”, we designed formative assessment lessons specifically to address conceptual understanding or strategic problem solving, complementing the procedural curriculum in many schools and showing how higher-level thinking may be taught and assessed. These lessons are now in use by millions of teachers and students across the US; evaluation shows remarkable gains in student learning (Herman et al., 2014).

Developing Design

In this section we shall outline some projects that led us to develop specific design principles and tactics—principles that continue to inform our work. Other projects, outlined in emma.mathshell.org, will be referred to as they arise in what follows.

Testing Strategic Skills (TSS 1980–88) developed a new model of examination-driven gradual change. The stimulus in this collaboration with England’s largest examination provider was our pointing out that, of the board’s list of 7 “knowledge and abilities to be tested” in mathematics, only 3 were assessed in the actual examinations. The board agreed to a novel strategic design with the following features: introduce one new task type each year, with 2 years notice to schools; provide integrated support in the form of materials; remove from the exam

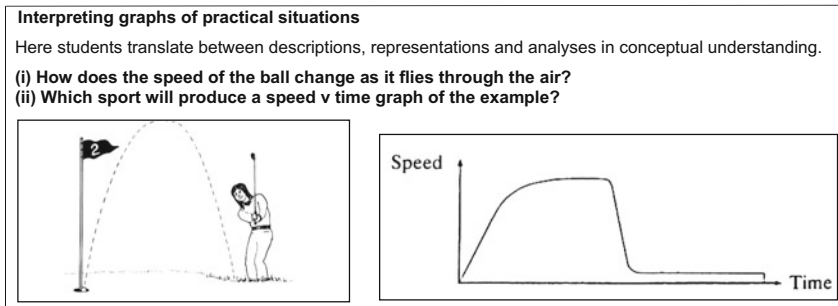


Fig. 3 From *The Language of Functions and Graphs*

syllabus 5% of the mathematical content to compensate. The Shell Centre developed the materials (Swan, Binns, Gillespie, Burkhardt, & The Shell Centre Team, 1987–89; Swan, Pitts, Fraser, Burkhardt, & The Shell Centre Team, 1985, 1984), which comprised 5 exemplar examination tasks,¹ lesson materials for 3 weeks teaching, along with materials for in-school professional development including video and software. The long-term goal was *to move year-by-year towards “tests worth teaching to”*—a target that still remains elusive worldwide. The gradual change model was popular with teachers, students and the board. Two modules were developed, one on problem solving, the other on concept development. *Problems with Patterns and Numbers* (Swan et al., 1984, see Fig. 2b) is concerned with generalization of mathematical situations. *The Language of Functions and Graphs* (Swan et al., 1985, Fig. 3) involves students in translating between representations of practical situations.

A national reorganization of testing ended this promising innovation—as reorganization so often does. ‘Replacement unit’ models have been used in many places, but the digestible pace of change and coherent well-engineered assessment and support of “TSS” remain rare.

This work led to a collaboration with Alan Schoenfeld’s group at Berkeley and others in the US in a series of projects on assessment and large-scale change. This still forms a major strand of our work, some of it described in the sections that follow.

Investigations on Teaching with Microcomputers as an Aid (ITMA 1980–88)
This project explored the potential of a single microcomputer with a large monitor in supporting the teaching of non-routine problem solving. Led by Rosemary Fraser and Richard Phillips, the design was largely based on “software microworlds” that stimulate investigation (Fig. 4).

Though inevitably “off-line” for the students, the approach proved powerful in various ways. We shall mention just one: “role shifting”. A study of 17 classrooms

¹For expert tasks, it is essential to show the *variety* that can be expected. We used the rubric “The following sample of questions gives an indication of the variety likely to occur in the examination”.

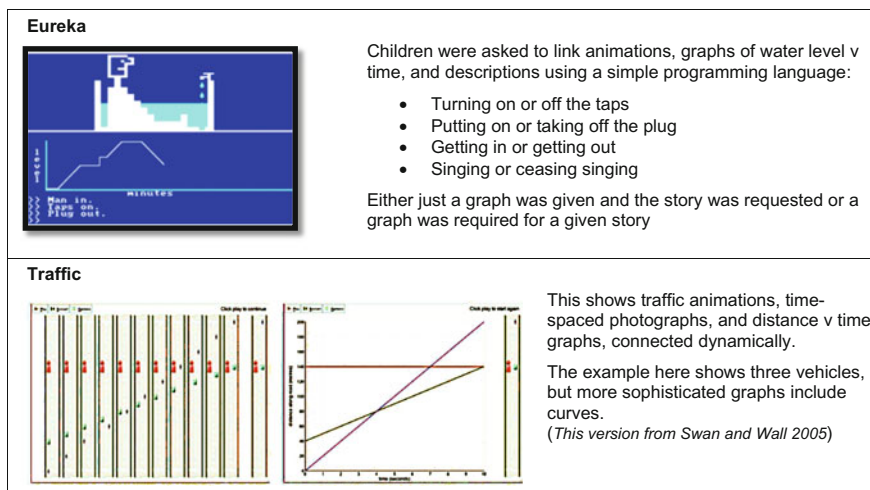


Fig. 4 ITMA microworlds

using ITMA software lessons (Burkhardt et al., 1988) developed a “roles analysis” which showed that the teachers² naturally moved from the traditional directive roles (called *manager*, *explainer*, *task setter*) into facilitative roles (*counsellor*, *fellow student*, *resource*). Students became *explainers* and *task setters*. Designing for role shifting has proved a powerful design tactic in our subsequent work: higher-level discussion and learning happen reliably when students move into teacher roles.

This strand of technology-based work has continued to inform our design more generally. For the UK Government’s **World Class Arena** (1999–2005), Daniel Pead led the development of tests of “Problem solving in Mathematics, Science and Technology” that were ‘computer + paper’ based. This (expensive) format reflected the limitations of AI—still weak after 50 years—in interpreting autonomous student reasoning. This work led to an analysis (see ISDDE, 2012) of the strengths and weaknesses of computers in the five essential aspects of summative assessment: presenting the task (strong); providing a natural working environment for the student (strong for text only subjects, weak for mathematics with its sketches and equations); capturing student responses (fine for text; for mathematics, only for novice tasks—or by scanning written responses); assessing responses (very weak for complex tasks); collecting and reporting scores (very strong).

Diagnostic Teaching (1983–2006) was the guiding principle for a linked sequence of small-scale studies, initially led by Alan Bell. It is based on students revealing their misconceptions through carefully designed “cognitive conflict” situations, then “debugging” them through group and class discussion (Bell, 1993; Swan, 2006). The outcome of this approach to formative assessment is improved

²With one exception: He loved the lessons but continued to stand by the screen and teach from the front.

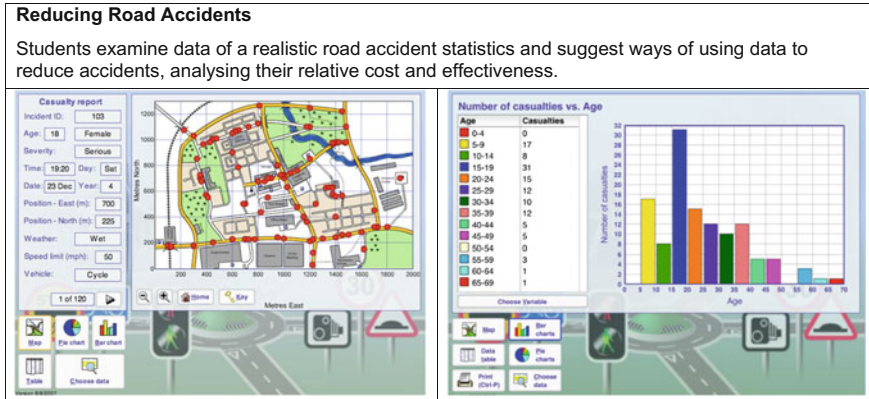


Fig. 5 From Bowland Maths: <http://bowlandmaths.org.uk>

long-term learning. Since Section “[Developing Conceptual Understanding and Logical Reasoning](#)” will describe this ongoing work, we will just point to the research strategy here.

Many studies in education are small-scale investigations of a specific *treatment*—a new approach to teaching a topic, perhaps studied in a few classrooms. If well done, such studies may reveal *trustworthy* insights about that system but without any evidence of *generalizability* (Schoenfeld, 2002). To provide evidence on design *principles*, the same research approach needs to be studied across a range of topics, designers, and teachers—as well as students. That has been the strategic design of the sequence of diagnostic teaching studies; the principles have proved robust.

Bowland Maths (2006–10) is a happier story. The goal was to develop 4-lesson units (“case studies”) for 14-year old students that showed the power of mathematics through real (or fantasy) world situations. We developed two of these. *How risky is life?* confronts students with the mismatch between popular ideas of various dangers and the data. In *Reducing road accidents* the students explore a graphical database of detailed accident report data (Fig. 5) in order to advise a town council on what safety measures to take. At a time when the learning goals of the national curriculum are broadening, this enrichment model of change has had some impact. We also developed professional development support modules based on a novel “sandwich model”, which we describe in Section “[Tools for Supporting Systemic Change](#)”.

Developing Conceptual Understanding and Logical Reasoning

In design we use principles from research on learning, notably that students learn through active processing: discussion and reflection in a social classroom leading to the internalization and reorganization of experience. This we developed into the following design principles, which underlie much that we describe below:

- Activate pre-existing concepts and problem solving strategies
- Allow students time to build multiple connections
- Stimulate tension—cognitive conflict—to promote questioning, re-interpretation, reformulation and accommodation
- Devolve problems to students
- Focus on reasoning—not just answers
- Expect students to explain their Interpretations and chains of reasoning
- Include reflective periods of ‘stillness’, for examining alternative meanings and methods.

We have identified a number of *lesson genres* that contribute to conceptual understanding:

- Interpreting and translating representations—*What is another way of showing this?*
- Classifying, naming and defining objects—*What is the same and what is different?*
- Testing assertions and misconceptions and justifying conjectures—*Always, sometimes or never true?*
- Modifying problems; exploring structure—*What happens if I change this? How will it affect that?*

We will illustrate the first of these genres below with a lesson on percentage change.

Diagnostic Teaching Research and Development

The sequence of diagnostic teaching studies, over several mathematical topics and teaching unit designers, showed that this approach leads to more robust long-term learning than direct instruction. Swan (2006) analysed the average pre-test, post-test and delayed-test scores in those studies. In each case, the intensive discussion and argument among students yielded more substantial long-term learning than standard methods—either exposition or guided discovery led by teachers.

To provide evidence of the generalizability of this result, the same research approach was studied, linked across a range of conceptual topics, and applied by students in non-obvious situations. The principles have proved robust and used in

subsequent projects. **Improving the Learning of Mathematics** (Swan & Wall, 2005), for example, was a collaboration with the UK government's Standards Unit which produced curriculum development support materials. This "box" was distributed to all UK schools and colleges, receiving an enthusiastic response from practitioners, researchers, and government inspectors.

Formative Assessment

A large-scale review of research by Black and Wiliam (1998) showed that the use of formative assessment, when well done, leads to remarkable increases in student learning. Wiliam and Thompson (2007) defined it thus:

Formative assessment is students and teachers using evidence of learning to adapt teaching and learning to meet immediate needs, *minute-to-minute and day-by-day*.

This is, of course, the essence of the Diagnostic Teaching approach. However, making formative assessment central to one's practice is a major departure from the "demonstrate and practice" form of pedagogy that lies at the core of most mathematics teaching, so it is extremely challenging for teachers. Our earlier work led the Bill & Melinda Gates Foundation to invite us to design lesson materials that enable teachers to acquire this expertise.

In the **Mathematics Assessment Project** (2010–14), working with our US partners, we designed 100 "Classroom Challenges"—20 formative assessment lessons for each grade 6 through 10—and refined them on the basis of structured observer reports through two rounds of trialing in US classrooms. Two thirds of the lessons are concept development focused, the others problem solving focused. We describe their design in more detail in Swan and Burkhardt (2014). There have been over 6,000,000 lesson downloads so far from the project website map.mathshell.org alone.

The Design of Concept Development Lessons

These lessons are designed with three complementary objectives:

1. **to reveal to the teacher, and the student, each student's current understanding and misunderstandings** of the central concept—as in all well designed diagnostic assessment.
2. **to move the student's understanding forward by a process of "debugging through discussion,"** in pairs and with the class as a whole—thus integrating diagnosis and treatment. This is crucial: diagnosis alone faces the teacher, again and again, with the considerable design challenge: "What shall I do to help this student?" A common response is to reteach the topic, but faster; it is not

surprising that this rarely helps. The diagnostic teaching approach used in the design of the Classroom Challenges reflects the observation (VanLehn & Ball, 1991) that the key characteristic of successful students is not that they remember procedures precisely but that they can detect and correct their own errors. “Debugging through discussion” develops that (higher level) skill. This more robust, long-term understanding reduces the time needed when re-visiting topics in later years.

3. **to build connections between different conceptual strands.** Mathematics content is best understood as a connected network of concepts and skills—as in other networks, the connections reinforce each other. The linear sequence of lesson-by-lesson teaching naturally develops “strands of learning”—strands that, for most students, have weaknesses, and often breaks, in them. Learning should involve active processing, linking new inputs to the student’s existing cognitive structure. Novice tasks alone produce fragmentation; rich tasks help to develop connections.

The design has the following sequence of activities:

- Expose and explore students’ existing ideas—*pull back the rug*
- Confront them with their contradictions—*provoke ‘cognitive conflict’*
- Resolve conflict through discussion—*allow time for formulation of new concepts*
- Generalize, extend and link learning—*connect to new contexts.*

“Increasing and decreasing quantities by a percent” (<http://map.mathshell.org/lessons.php?unit=7100&collection=8>) is a lesson that shows how this works. It is designed to enable students to detect and correct their own and each other’s misconceptions in this often-challenging topic—and to build connections.

During a prior lesson, a sheet of tasks on percent changes is given to students. It includes:

In a sale, all prices in a shop were decreased by 20%. After the sale they were all increased by 20%. What was the overall effect on the shop prices? Explain how you know.

The vast majority of students (and many adults!) think there is no overall change. $Price - 20\% + 20\% = Price$. You just add % changes.

Real understanding involves knowing that we are combining *multipliers*:

$Price \times 0.8 \times 1.20 = Price \times 0.96$ —a 4% reduction.

These are challenging ideas to get across by explanation, or by standard ‘demonstrate and practice’—the prevalence of the misconception makes that clear. However, the challenges of understanding percent increases and decreases become much more accessible when students confront them in their own work, as in this lesson.

After the diagnostic pre-assessment, the students are given four cards with carefully chosen numbers (100, 150, 200, 160) for the corners of a poster, and ten arrow cards. Eight of the arrow cards contain expressions like “increase by 50%” or “decrease by 25%”; two are blank. The students’ task, in pairs or threes, is to place

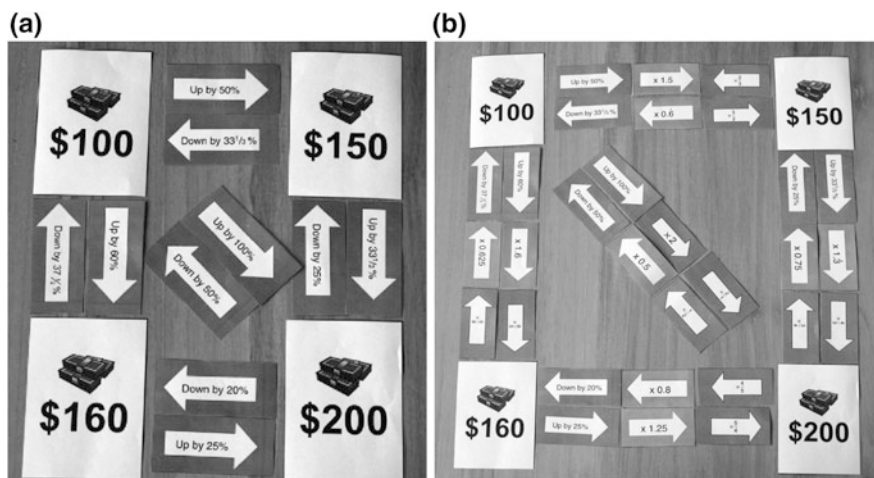


Fig. 6 a The first poster, with “up and down ...”. b The second poster, with “multiply by ...”

the arrow cards that correctly indicate the relationships between the 4 numbers. To begin with, standard misconceptions appear: placing the “increase by 50%” arrow between 100 and 150 is straightforward, but many students place the “obvious” reverse, “decrease by 50%” alongside it. Then they discover they need that arrow to connect 200 to 100. This provokes discussion, and provides room for questions from the teacher. Teachers are prompted to ask students to clarify their thinking and share ideas with other students. The result is Fig. 6a.

Then two further sets of arrow cards are distributed, and placed in a similar way: first multiplications by decimals (“ $\times 1.5$,” etc., Calculators are also given out at this point.) and then multiplication by fractions (“ $\times 3/2$,” with “ $\times 2/3$ ” for the inverse—a key insight for proportional reasoning). This, shown in Fig. 6b, exemplifies Objective 3 above, linking topics that are initially taught separately. (Linking the two numbers 150 and 160 is kept in reserve, for students who move rapidly through the lesson.) Although aimed at Grade 7, this lesson also provides valuable “stress testing” of understanding in later grades. It shows the importance of building connections, and the way discussion on rich tasks drives this.

This lesson exemplifies a broader design goal: to help students see results from different perspectives. Richard Feynman³ put it thus: “If you find a result one way, it is worth thinking about. If you can show it in two ways, it may well be true. If you can show it three ways, it probably is.” Not proof, but deep understanding.

³Private communication to HB, then working in physics at Caltech.

Developing Strategies for Problem Solving

A problem is a task that the individual wants to achieve, and for which he or she does not have access to a straightforward means of solution. (Schoenfeld, 1985)

The ability to tackle such problems is the essence of mathematical expertise. So helping teachers to teach problem solving effectively has been an ongoing strand of our work, starting with *Problems with Patterns and Numbers*. Here we shall describe the most recent design: the formative assessment lessons in problem solving from the Mathematics Assessment Project.

Structure of a Problem Solving Lesson

These have a rather different design from the concept-development lessons, though the time-structure is similar. They use, instead of alternative (mis)conceptions, alternative solutions to a single rich problem around which the lesson is built. This is the sequence:

- In a prior lesson the students spend around 20 min tackling an expert task, individually and unaided.
- Before the main lesson the teacher assesses the work, looking for different approaches and, with guidance from the “Common Issues” table, prepares qualitative feedback in the form of questions—sometimes individually but often for the class.
- In the main lesson the students review their own work in the light of the teacher’s feedback and write responses.
- Collaborative work in pairs or threes follows, with students working to share ideas and to produce joint solutions.
- Carefully chosen examples of other students’ work using different mathematical methods are introduced. The groups are asked to review and critique the various solutions, in their groups, then as a whole class discussion.
- Whole class discussion focuses on the payoff of mathematics at different levels. (The sample student work allows us to show more powerful mathematics than is likely to arise in a typical class.)
- Students improve their solutions to the initial problem, or one much like it.
- Finally, in a period of individual reflection, students write about what they have learned.

We will illustrate this design by sketching two examples. As always, there is no substitute for reviewing the complete lesson guide and, if possible, trying the lesson yourself.

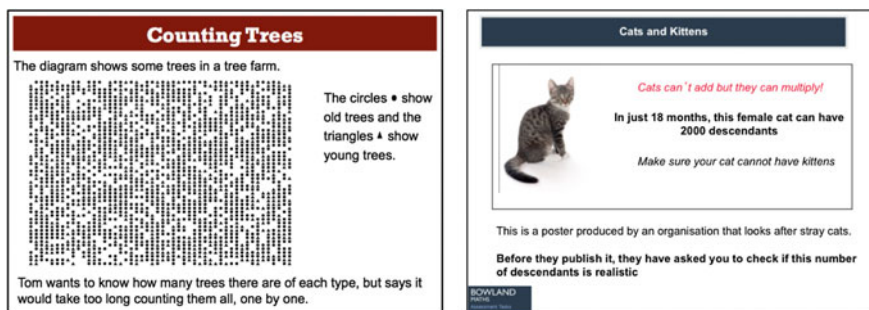


Fig. 7 The tasks from *Counting Trees* and *Cats and Kittens*

Problem Solving Tasks: Counting Trees and Cats and Kittens

The problem-solving-focused Classroom Challenges make fundamental use of core content but do so in the context of challenging students to use the mathematics in ways that call for being strategic and logical. They emphasize working through and explaining their problem solving processes.

Students are also expected to analyse alternative solutions, each incomplete or incorrect, in order to compare and challenge their approaches. These will stimulate further analysis and development. Consider the “Counting Trees” task in Fig. 7. The fundamental decisions in working on this task are strategic. If you don’t count all the trees, then, you need to sample. But how should you do so? How large a sample; where do you take it from; how do you scale up? Do you want to get more sophisticated, and take a few samples, and average results? Of course, the more work one does, the more accurate the results are likely to be—but the less effort one saves.

In this lesson fundamental strategic (and mathematical) considerations emerge as different groups compare their approaches. Furthermore, all of this work is grounded in applications of ratio and proportion: proportionality underlies the notion of sampling and scaling. Hence this problem-solving lesson, like the concept-focused lessons, engages students in linking a range of fundamental concepts.

In the “Cats and Kittens” lesson, an advertisement advising neutering says a female cat can have 2000 descendants in 18 months. It gives the following data: *A female cat can get pregnant at age about 4 months. The pregnancy lasts about 2 months. A typical litter is 4–6 kittens. A cat can have about 3 litters a year, until they are 10 years old.*

The students are asked to work out whether the estimate of 2000 is reasonable? This much more open problem involves overlapping exponential growth but, for middle school students, the essential challenge is finding and using an appropriate representation to organize the calculations. (Student work is shown in emma.-mathshell.org, where there are links to the complete lessons)

The lessons from the Mathematics Assessment Project have proved popular with US teachers—and independent evaluators. On student learning, a report on 9th

grade algebra students (Herman et al., 2014) found the average gain in algebra after 8–12 days of instruction using Classroom Challenges was 4.6 months more than the norm for grade 9. How could that be? There are a number of explanations. In content terms, these lessons are *synthetic*: they pull together prior learning and enhance it. But the pedagogy of the lessons, and its impact on the teachers' style, is at least as important.

Tools for Supporting Systemic Change

Many groups around the world now know how to enable typical teachers to teach much better mathematics much more effectively. Nobody knows how to lead school systems to make the changes needed for this to happen on a large scale.

We believe that this is the central challenge of our time. We believe that all the key players (policy makers, the research community, administrators, principals and teachers) play a role in this systemic failure and must be part of the resolution. In this final section, we look at what barriers seem to impede improvement, and how we might help to overcome them (Burkhardt & Schoenfeld, 2003; Burkhardt, 2009, 2015). Despite the limited success so far, we remain hopeful.

With our US partners we have recently taken on a specific challenge: *Can we develop effective system-level tools?* While teaching materials and assessment tools are well-recognized as important, and professional development tools are slowly being accepted, people in leadership roles in school systems (local school districts, states, nations) have not seen the value that tools could provide for them. Working with 10 school system partners in a Mathematics Network of Improvement Communities (NIC), our experience so far shows that we can develop tools to meet specific challenges that the partners specify as important, and that these tools can be helpful. We mention a few in what follows. But it is still early days.

Tools for Professional Development

We start with an area of success. Everyone recognizes that teaching quality is crucial to student learning, that improvement involves qualitatively new challenges for most teachers of mathematics of the kind discussed above, and that ongoing support for teachers in developing their professional expertise is essential. While we have shown that teaching materials can provide powerful support for teachers, they need to be complemented by effective professional development (PD) programs. To implement that on a large scale will require well-engineered tools.

Why do we need materials for professional development? First there is a mismatch of scale: the number of PD leaders with the right expertise in this area is far too few for the number of teachers who need support. Secondly, developing PD that *actually leads to changes in teachers' classroom practice*, and is cost-effective in

teacher time, is a challenging design problem. (Evaluations of PD habitually only ask if the participants found it valuable—a very different criterion.)

We have found two key design features for effective professional development. It is activity-based, since active learning by processing issues is as important for teachers as it is for students. It is on-going, since high-quality teaching is the product of decade-long professional learning within a strong theoretical framework like TRU, Teaching for Robust Understanding (Schoenfeld, 2014). This design challenge implies a need for well-engineered materials. Our approach to PD has always been based on teachers learning ‘constructively’ from carefully designed experiences in their own classrooms. In our “sandwich model” (Swan & Pead, 2008) a group of teachers first meet for a structured discussion of a key issue of pedagogy: ‘Handling classroom discussion in a non-directive way’, for example. They together prepare a lesson based around teaching materials we design. They each teach the lesson in their own classrooms, or observe each others’ lesson, then prepare feedback for the second session together. In this discussion they report back and return to a structured reconsideration of the issues and of the next step in their development.

These modules support the first stages of a route towards our longer-term goal of helping teachers to become part of a professional learning community, using Japanese ‘Lesson Study’ as a model. In NIC we have developed tools for system leadership on approaches to the design of PD, and of lesson study. The *NIC Classroom observation tool* is designed for use by school principals and others who, despite not having a mathematics background, are required by their systems to observe and evaluate mathematics teachers. Based on the TRU framework, this tool is designed to help ‘non-math-ed people’ pick out the important things in the classroom, focusing on the nature and quality of what students are asked to do and how they are responding to it. (“A quiet class, working hard” may impress but it is not the core indicator of a good learning environment.)

Strategic Design Opportunities

In looking for ways to overcome systemic obstacles, it’s worth looking for ‘leverage’ points that offer a way to answer the fundamental question facing reform: “*Why should they change?*” Here we list a few responses, starting with those where relatively small changes can have big effects.

Design ‘tests worth teaching to’. Though high-stakes examinations are often barriers to progress, they can be and have been powerful levers for improvement (Burkhardt, 2009 gives examples, including TSS). The empirical fact that *What You Test Is What You Get* in most classrooms means that better tests can lead to lessons of higher-quality, as long as teachers are given effective support in meeting the new challenges such tests present. This needs an explicitly specified balance and weighting across the elements in Table 1: factual knowledge and procedural fluency, conceptual understanding and logical reasoning, and problem solving and

strategic competence—with comparable weightings in both curriculum and assessment.

Facts and procedures are usually dominant in tests because they are simplest to assess (and to teach). Assessing conceptual understanding is more complex, since it involves chains of autonomous student reasoning. Problem solving also requires extended reasoning, including choices of suitable mathematical tools and their subsequent application. Assessing these needs different design tactics as well as richer tasks: for example, asking students to critique sample responses to a complex task. The design of this broader and deeper kind of assessment depends on task designers with a wider range of skills than test providers have needed for ‘novice tests’. The educational disasters that have so often been produced in the process of turning (usually well meaning) intentions into actual high-stakes tests (Burkhardt, 2009) make this a crucial opportunity for progress.

Aim for alignment across curriculum standards, teaching materials, assessment and professional development. This avoids sending mixed signals to teachers. NIC has developed a *Program Coherence Health Check* tool based on comparing the balance of task-types in the various aspects of the system’s improvement program—and describing options for improving the alignment.

Plan the pace of change. Politicians try to “fix the teaching mathematics problem” in ways that they wouldn’t try in other fields, for example medicine, where gradual improvement is accepted as inevitable. Well-engineered gradual change can work in education, too, while politics-driven “Big Bang” methods typically yield only superficial change. The appropriate strategic design question is “How big a change can teachers carry through effectively, year-by-year, given the support we can make available?” We have observed in Japan and other countries that deep challenges to teacher expertise can, if done well, be exciting for teachers. Developing such long-term professional development practices is vital.

Structural Design Tactics

Moving from the strategic to the structural, the following have proved powerful design tactics.

Use replacement units to support gradual change at a digestible pace. TSS modules (Fig. 3) provide coherent support, integrating assessment, curriculum and professional development materials. Software microworlds, as in ITMA (Fig. 4), help teachers handle inquiry-based learning, with teachers and students naturally shifting roles. “Classroom Challenges” have proved powerful in advancing student learning. Replacement units like these can provide ‘protein supplements to a carbohydrate curriculum’.

Use exemplars Descriptions alone tend to be interpreted within the reader’s prior experience. We hope the figures in this paper, and in emma.mathshell.org, show the value of task and lesson exemplars in communicating meaning.

Identify target groups be they students, teachers, PD leaders, superintendents, and/or policy makers, and co-develop your tools with them. Who do we need this to work for? Not just the enthusiasts. We found that “second worst teacher in your department” works well with designers as a teacher target group!

Distribute design load “How much guidance shall we give to teachers?” is a key design question. Too little and they won’t have enough support; too much and they won’t read it. We offer detailed guidance when we are better placed to do so than the teachers we serve. (The ‘trials teachers’ usually suggest more.)

Design and Development Tactics—and Costs

The following tactics help to make the Shell Centre approach cost-effective.

- **“Fail fast, fail often”**—rapid prototyping with quick feedback allows the design team to learn quickly.
- **Make feedback cost-effective** by getting rich feedback from small samples. We find 3–5 classrooms is large enough to distinguish general from idiosyncratic features, and small enough to allow the rich observational data needed to inform revision.
- **“Design control”** describes our identifying who, after discussions within the team, will take the design decisions in each area of design. The alternative, seeking consensus, is too expensive in time and doesn’t always produce good designs.

Research-based design and development is much more expensive than traditional “authoring”—for us, typically US\$3000 per task, \$30,000 per lesson. But good engineering can ensure that the activities work well and that the materials communicate, enabling typical users to meet ambitious educational goals. Surprisingly, though these sums look large, the cost of using this approach for the whole curriculum would still be negligible within the cost of running a large education system.

The Case for “Big Education”

Other fields accept that big problems in complex systems need big coherent collaborations using agreed common methods and tools, specifically developed for key problems of practice. The CERN Large Hadron Collider and the Human Genome Project are two obvious examples. Most medical research is of this kind. We argue (Burkhardt, 2015) that research in education needs a similar approach if, for example, we want the better evidence on the generalizability of research results that

design needs. This is a challenge for a field whose academic value system has encouraged new ideas over reliable research, new results over replication and extension, personal research over team research, disputation over consensus building, academic papers over products and processes—all of which conflict with the goal of well-founded large-scale impact on practice.

.... and finally

The work we have described here is the product of the brilliant individuals we have been fortunate enough to work with over the last 35 years: our colleagues in the Shell Centre team—Alan Bell, Rosemary Fraser, Richard Phillips, Daniel Pead, Rita Crust—and many other researcher-designers in Nottingham and around the world, notably Alan Schoenfeld, Phil Daro, David Foster and the Silicon Valley Math Initiative, Sandra Wilcox and her Michigan State team, Kaye Stacey and other outstanding Australians, and the ITMA team at Marjons. In addition, enormous thanks are due to the teachers in whose classrooms these tools have been trialed and observed. The work has been supported by a variety of willing-to-be guided funding agencies from government, assessment, and the foundations in the UK, the US and the EU.

Finally, we look at the issues that the team has faced over the last 40 years in simply surviving, when so many fine design teams have struggled, often disappearing into other work. The account in this chapter, of coherent strands of research and development over decades, shows that longevity is important; with funding uncertain from project to project, it is not easy to achieve. We have found strategies that can help. First, it is important to diversify the sources of funding—each funding agency has priorities which change over time. For example, in the 1980s following the Cockcroft Report (1982), the UK Government saw the need for R&D to help the system meet the new goals but, with the 1989 introduction of the National Curriculum, government saw implementation as its only concern. However, at this time in the US the NCTM Standards (NCTM, 1989) appeared, which led to a surge of support for R&D over the next 15 years from the National Science Foundation. This pattern, continued across the US, UK and the European Union, along with some luck in the timing, has helped our team survive. This illustrates the second element of strategy—to build long-term relationships across the mathematics education world. We all benefit from the mutual enrichment and support in many ways, including funding. Last but not least, it is strategically important to work on projects that you think are important, looking for the overlap of funding opportunities with challenges that seem to have promise for moving the field forward. To this end, we have been proactive in proposal design—funders rarely understand what they want in any depth and, we have found, are happy to let you convince them to modify their original ideas. These strategies may, with luck, give a dedicated team the time to become good at the deep engineering research that yields products that combine educational ambition with substantial impact on practice—the essence of the Emma Castelnuovo criteria.

References

- Bell, A. (1993). Some experiments in diagnostic teaching. *Educational Studies in Mathematics*, 24 (1).
- Black, P. J., & Wiliam, D. (1998). *Assessment and classroom learning assessment in education*, 5, 7–74.
- 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. (2009). On strategic design. *Educational Designer*, 1(3). Retrieved from: <http://www.educationaldesigner.org/ed/volume1/issue3/article9>
- Burkhardt, H. (2015). Mathematics education research: a strategic view. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd ed.). London: Taylor and Francis.
- Burkhardt, H., Fraser, R., Coupland, J., Phillips, R., Pimm, D., & Ridgway, J. (1988). Learning activities & classroom roles with and without the microcomputer. *Journal of Mathematical Behavior*, 6, 305–338.
- Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: towards a more useful, more influential and better funded enterprise. *Educational Researcher*, 32, 3–14.
- Cockcroft Report. (1982). *Mathematics counts, report to the UK government of the committee of enquiry*. London, UK: HMSO.
- Herman, J., Epstein, S., Leon, S., La Torre Matrondola, D., Reber, S., & Choi, K. (2014). In *Implementation and effects of LDC and MDC in Kentucky districts (CRESST Policy Brief No. 13)*, National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles: University of California.
- ISDDE (2012). Black, P., Burkhardt, H., Daro, P., Jones, I., Lappan, G., Pead, D., et al. (2012). High-stakes examinations to support policy. *Educational Designer*, 2(5). Retrieved from <http://www.educationaldesigner.org/ed/volume2/issue5/article16>
- NCTM. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Polya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.
- 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.
- Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? *Educational Researcher*, 43(8), 404–412. doi:10.3102/0013189X1455
- Swan, M. (2006). *Collaborative learning in mathematics: A challenge to our beliefs and practices*. London: National Institute for Advanced and Continuing Education (NIACE) for the National Research and Development Centre for Adult Literacy and Numeracy (NRDC).
- Swan, M., Binns, B., Gillespie, J., Burkhardt, H., & The Shell Centre Team. (1987–89). In *Numeracy through problem solving: Five modules for teaching and assessment: Design a board game, produce a quiz show, plan a trip, be a paper engineer. Be a shrewd chooser*. Harlow, UK: Longman. Downloadable from <http://www.mathshell.com>
- Swan, M., & Burkhardt, H. (2014). Lesson design for formative assessment. *Educational Designer*, 2(7). Retrieved from: <http://www.educationaldesigner.org/ed/volume2/issue7/article24>
- Swan, M., Pitts, J., Fraser, R., & Burkhardt, H., The Shell Centre Team. (1984). In *Problems with patterns and numbers*. Manchester, U.K.: Joint Matriculation Board. Downloadable from <http://www.mathshell.com>
- Swan, M., Pitts, J., Fraser, R., Burkhardt, H., & The Shell Centre Team (1985). In *The Language of Functions and Graphs*. Manchester, U.K.: Joint Matriculation Board. Downloadable from <http://www.mathshell.com>

- Swan, M., & Pead, D. (2008). Bowland maths professional development resources. Online: <http://www.bowlandmaths.org.uk/pd>. Bowland Trust/Department for Children, Schools and Families.
- Swan, M., & Wall, S. (2005). In *Improving the learning of mathematics*. UK: Department for Education and Skills.
- VanLehn, K., & Ball, W. (1991). Goal Reconstruction: How Teton blends situated action and planned action. In K. VanLehn (Ed.), *Architectures for Intelligence* (pp. 147–188). Hillsdale, NJ: Erlbaum.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Open Access Except where otherwise noted, this chapter is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

