Margolinas, C. (Ed.). (2013). Task Design in Mathematics Education. Proceedings of ICMI Study 22 . Oxford.

Task design for systemic improvement: principles and frameworks

Hugh Burkhardt
Shell Centre Team, Centre for Research in Mathematics Education,
University of Nottingham and UC Berkeley.

Malcolm Swan
Shell Centre Team, Centre for Research in Mathematics Education,
University of Nottingham and UC Berkeley.

Tasks play many roles in mathematics education, some of them unintentional. They come in many forms, spanning many dimensions, including those of mathematical content, processes and practices, length, and modes of working. In this paper we point out the crucial role that tasks play in forwarding or preventing the process of improvement of teaching and learning – and thus of an education system. We argue that multi-dimensional schemes of task classification have a powerful role to play in the design of tasks and task sequences.

Keywords : classification, dimensions, balance, curriculum, assessment, sequences

Tasks in this paper will mean "activities that students are asked to do, in which mathematics has an important role". Students meet tasks in the course of classroom learning, in informal tests or examinations, in the course of their everyday lives, and in other school subjects. Task design for all these roles has been a central concern of the work of Shell Centre team over the last 30 years, which the authors have led (Swan and Burkhardt, 2012). Our mathematical and epistemological perspectives are eclectic, reflecting the priority in the team's research given to direct impact on systemic improvement.

In discussing tasks and their design, exemplification is absolutely essential to clarify the meaning of descriptions. As so often with print, the limitations of length in this paper mean that we cannot give examples here; we shall accept the opportunity to extend the range on-line, with links in the text to examples of various kinds.

1. Tasks and their various roles

The major focus of this Study will be on the nature and use of tasks in teaching and learning in classrooms. From the systemic viewpoint of this paper, however, it is appropriate to describe and discuss their roles starting with the system, and moving through assessment in its various forms to the classroom.

In specifying a curriculum Currriculum[46] specifications are normally in the form of descriptions of principles, amplified by an analytic model of the domain. The latter in particular, vary greatly in length, from a few pages on competencies in the Denmark to long and detailed lists of mathematical content in British or US documents. The one common feature is that they are largely expressed through language. Such descriptions do not, in fact, specify learning or performance goals – for example, lists of mathematical content could be taught and assessed entirely through short tasks on separate elements, or through substantial projects in which th student chooses and uses appropriate elements of content and process for th investigation in hand – or, more sensibly, for a balanced variety of types of performance. Task exemplars can play a crucial role in reducing this ambiguity. We have argued (Burkhardt, 1990) that a curriculum specification needs three different elements: an analytic model of the domain: an exemplar task set, with each task linked to the model; a list of the range of classroom learning activities that should be involved (for a brief example, see Cockcroft Report, 1982, paragraph 243).

A curriculum
spec needs 3
KEY
ELEMENTS:

(1) ANALYTIC
MODEL OF THE
DOMAIN;

(2) AN
EXEMPLAR
TASK SET w/
links to the model

(3) A LIST OF
RELEVANT
CLASSROOM

In high-stakes examinations In countries that have tests where the results have life consequences, the range and balance of types of task in the tests have a strong influence on the range and balance of classroom learning activities (see e.g. OFSTED 2012). Indeed they often seem to define the *de facto* "implemented curriculum" in most classrooms, whatever the intended curriculum of the last paragraph may say. So high-stakes assessment, and the tasks the tests contain, plays three roles:

- A: to 'measure' performance – ie
  "*to enable students to show what they know, understand and can do*"

but also, with high-stakes assessment that impacts students' and teachers' lives, inevitably

- B: to exemplify the performance goals – assessment tasks communicate vividly to teachers, students and their parents what is valued by society, and thus
- C: to drive classroom learning activities *(What You Test Is What You Get)*

These roles carry responsibilities for test designers and those who commission tests – responsibilities that are widely ignored. Psychometricians, too, focus on measurement and statistical error, ignoring the systematic error that comes from assessing only a part of what you want students to learn. Ignoring roles B and C, and the systemic responsibility for test design that they imply, is a major source of the mismatch between intentions and outcomes in school systems.

In classroom assessment In some countries including ours, classroom assessment has traditionally reflected the formal tests, with similar task sets. This is a natural way for teachers to check progress towards an important goal. Some teachers, aware of the limitations of the tests, have always used a broader range of task types in the classroom. In the last decade there has been growing awareness of the power of formative assessment, when well done, in forwarding student learning (see Black and Wiliam 1998, 2001). This approach integrates assessment and teaching in a form where the design of *task sequences* plays a crucial role. This is challenging for teachers so there has been work on the design of support, initially through live

---

[46] By curriculum, we mean the whole set of learning activities that a student experiences in school.

professional development and, more recently, through classroom materials (Swan et al., 2011)

In teaching and learning We expect the roles of tasks in teaching and learning to be the focus of most of the papers in this study – so here we shall be brief. That their range and variety should cover all the learning and performance goals of the intended curriculum is clear. What makes this area rich are the issues of, and principles for, designing *task sequences* that will lead students along the road to the understanding and performance goals. One may view task sequences as the spine around which all teaching is built, whether they be the succession of closely related exercises of "incremental learning" and its behaviourist relatives or, at the other extreme, the mathematical microworlds of "open investigation", where creating a task sequence by posing questions is part of the student's responsibility. We shall say something more in this in Section 3.

Tasks should cover all learning & perf goals

We'll focus on designing
TASK SEQUENCES here

2. Task difficulty

The issue of task difficulty is often ignored but is important in all aspects of task classification and design. 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 task complexity that affect the difficulty it presents.
- *unfamiliarity* – non-routine tasks (those which aren't just like the tasks one has practiced solving) are more difficult than routine exercises.
- *technical demand* – tasks that require more sophisticated mathematics for their solution are more difficult than those 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, these factors 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 employs a routine skill. Focusing on technical aspects alone can lead to rich tasks as being dismissed as "below grade".

The difficulty of a task is determined by trialling the task with a random sample of students drawn from the target population. All assessment tasks, whether for use in the classroom or in summative tests, should be developed in this way, establishing their level of difficulty without undermining their validity as good mathematics.

3. Task variety and task classification

Given the range of roles that tasks play, outlined above, it is clear that appropriate forms of task classification may be useful for various purposes. In this section we set out some schemes and ways they have proved useful in supporting various task roles. There is a constructive duality between holistic and analytic dimensions of classification.

Novice, Apprentice and Expert tasks

This simple holistic dimension of classification (Swan et al 2011) has proved useful in drawing attention to the mismatch between widely accepted goals of mathematics education and current practice in both assessment and curriculum[47].

Mathematical skills and practices can be taught and/or assessed partly in isolation, partly under scaffolded conditions, and partly when students face substantial problems without scaffolded support. We call tasks that assess these three different types of performance *novice*, *apprentice*, and *expert* tasks respectively. More specifically:

- *Expert Tasks*. Experts solve problems as they arise. Expert tasks are rich tasks, each presented in a form in which it might naturally arise in mathematics, science or daily life. They require the effective use of problem solving strategies, as well as concepts and skills. Performance on these tasks indicates how well a person will be able to do and to use mathematics beyond the mathematics classroom. Expertise is the end goal of mathematics education.
- *Novice Tasks*. Novices are learning the tools of the trade. Novice tasks are short items, each focused on a specific concept or skill. Reflecting the high-stakes assessment, mathematics teaching and learning in Britain and the US is mainly focused on novice tasks.
- *Apprentice Tasks*. Apprentices solve problems, but usually carefully structured problems with guidance from an expert. Apprentice tasks are substantial, often involving several aspect of mathematics, and structured so as to ensure that all students have access to the problem. Students are guided through a "ramp" of increasing challenge to enable them to show the levels of performance they have achieved. Because the structure guides the students, the strategic demands and the range of mathematical practices involved are at a comparatively modest level. Apprentice tasks have a role in developing expertise.

A Framework for Balance

Clearly, classification needs to go well beyond this. The NSF-funded project *Balanced Assessment for the Mathematics Curriculum* aimed to design assessment that reflected the goals set out in the NCTM Standards (NCTM, 1989, 2001). We set out two supervening design principles: *Curriculum balance*: a test should be such that a teacher who "teaches to the test" is led to deliver a curriculum balanced in accord with the Standards. *Curriculum value*: doing the assessment tasks should be a worthwhile learning experience. To articulate what this means we developed the *Framework for Balance*, summarised in the table below.

The headings are self-explanatory except, perhaps, for *reasoning length*. This is the time envisaged for the student to work on the longest prompted section of the task – so a 10-minute task that is structured into many equal parts may have a short reasoning length. (Driven by the naive criterion-referencing behind the National Curriculum, this is common in the UK, where tasks often consist of a sequence of

---

[47] The reader can find examples of each at http://map.mathshell.org.uk/materials/tasks.php

short items, set in a common context.). The most important features of the Framework for Balance are:

- the classification is multidimensional, addressing the major aspects of performance
- the dimensions are both analytic (content, process, etc.) and holistic (task type, openness, goals, etc.)
- it provides a method of choosing a set of tasks for a tests to meet conditions, particularly balance, for this
- the analytic dimensions are handled semi-quantitative, with the elements of content or process in a task given rough proportions
- an associated "balancing matrix" can be used to ensure that, while every combination of properties cannot be assessed, the main dimensions are samples with appropriate weight.

This approach was first used for balancing collections of classroom materials (Balanced Assessment 1997-99). It also produced a way around a design dilemma: the more constraints you impose on a task designer (such as a cell in a content matrix to assess), the poorer the holistic quality of the tasks that result. The alternative approach is to free designers to design good mathematical tasks, classifying them later and choosing a balanced set for each test.

Framework for Balance

Mathematical Content Dimension

- **Mathematical content** in each task will include some of:
  - **Number and Operations** including: number concepts, representations relationships and number systems; operations; computation and estimation.
  - **Algebra** including: patterns and generalization, relations and functions; functional relationships (including ratio and proportion); verbal, graphical tabular representation; symbolic representation; modeling and change.
  - Measurement including: measurable attributes and units; techniques tolls and formulas.
  - **Data Analysis and Probability** including: formulating questions, collecting, organizing, representing and displaying relevant data; statistical methods; inference and prediction; probability concepts and models.
  - **Geometry** including: shape, properties of shapes, relationships; spatial representation, location and movement; transformation and symmetry; visualization, spatial reasoning and modeling to solve problems.

Mathematical Process Dimension

- **Phases** of problem solving include some or all of:
  Modeling and Formulating;
  Transforming and Manipulating;
  Inferring and Drawing Conclusions;
  Checking and Evaluating;
  Reporting.
- **Processes** of problem solving, reasoning and proof, representation, connections and communication, together with the above phases will all be sampled.

Task Type Dimensions

- **Task Type** will be one of: design; plan; evaluation and recommendation; review and critique; non-routine problem; open investigation; re-presentation of information; practical estimation; definition of concept; technical exercise.
- **Non-routineness** in: context; mathematical aspects or results; mathematical connections.

Considerations
for Task Types

- open middle
- non-routine
- length
- modes of working, etc

- **Openness** –tasks may be: closed; open middle; open end with open questions.
- **Type of Goal** is one of: pure mathematics; illustrative application of the mathematics; applied power over a practical situation.
- **Reasoning Length** is the expected time for the longest section of the task.

Circumstances of Performance Dimensions

- **Task Length**: in these tests most tasks are in the range 5 to 15 minutes, supplemented with some short routine exercise items.
- **Modes of Presentation, Working and Response**: these tests will be written.

This idea has been taken further: Daro and Burkhardt (2012) proposed the development of a "population of tasks" that epitomises the curriculum goals, and from which tests will be drawn as balanced samples.

While we do not present these classification schemes as definitive (though they have worked well for specific purposes), we do see task classification as an important part of task design.

4. Principles for the design of tasks and task sequences

We have recently written in some detail[48] on the principles and processes of task design (Swan and Burkhardt, 2012). Here we have space for a bare list of principles. We argue that curriculum and assessment should be built on tasks that:

Principles for task design & sequencing

1. *Reflect the curriculum in a balanced way*. Assessment should be based on a balanced set of tasks that, together, provide students with opportunities to show *all* types of performance that the curriculum goals set out or imply.
2. *Have 'face validity'*. Assessment tasks should constitute worthwhile learning activities in their own right. The tasks should be recognizable as problems worth solving – because they are intriguing and/or potentially useful.
3. *Are fit for purpose*. The nature of the tasks and scoring should correspond to the purposes of the assessment. Individual tasks should assess students' ability to *integrate* as *mathematical practices* their fluency, knowledge, conceptual understanding, and problem solving strategies. These aspects should not be assessed separately.
4. *Are accessible yet challenging*. Tasks should be accessible with opportunities to demonstrate both modest and high levels of performance, so the full range of students can show what they can do (as evidenced by high response rates with a wide range of levels of response).
5. *Reward reasoning rather than results*. Tasks should elicit chains of reasoning, and cover the phases of problem solving (formulation, manipulation, interpretation, evaluation, communication) even though their entry may be scaffolded with short prompts to ensure access.
6. *Use authentic or 'pure' contexts*. Assessment should contain tasks that are 'outward-looking', making connections within mathematics, with other subjects, and to help one to better understand life and the outside world. As in the real world, they may contain insufficient data (where the student makes assumptions and estimates) or redundant data (where the student makes selections). Students may be asked to respond in a given role: e.g. a designer, planner, commentator, or evaluator. Tasks that use contrived contexts should be avoided.
7. *Provide opportunities for students to make decisions*. Tasks should be included that encourage students to select and choose their own methods, allowing them to

---

[48] See http://www.educationaldesigner.org/ed/volume2/issue5/article19

surprise or delight. Some may be open-ended, permitting a range of possible outcomes.
8. *Are transparent in their demands*. Students should be clear what kinds of response will be valued in the assessment.

Task sequences in curriculum design

Task sequencing

- incremental learning
- focus on concept development
- some will focus on interpretation of ideas
- some will focus on contrasting alternative approaches

- All should activate prior knowledge, get Ss to tackle a task, & prompt deeper reflection

More broadly, we see the concept of *task sequence* at the heart of curriculum design. Too large a topic for this paper with its systemic focus, it will be the focus of a forthcoming article. Here we will just reinforce the idea with a few examples.

We have already noted that the "incremental learning" approach to curriculum design, characterised by small steps, is strongly reinforced by the nature of assessment. Instead, we have cotinued developing an approach that is very different (see Swan, 2006).

We distinguish whether task sequences are designed primarily to foster conceptual development or problem solving processes. The focus of the first is on discussing different *interpretations* of mathematical ideas; the second is on the contrasting alternative *approaches* that may be taken. In both cases we begin by seeking to find out students prior knowledge, by asking them to tackle a carefully chosen task individually, unaided. Their responses are assessed by the teacher, outside the classroom, who must then prepare a series of questions (tasks) designed to prompt students' deeper reflection. We provide a set of questions matched to typical responses and to assist the teacher in this.

In a problem solving lesson, students are then invited respond to these questions and form small groups to produce joint solutions that both combine the best of their individual ideas, and that address the teachers' questions. A sharing of alternative approaches is then undertaken, akin to the Japanese practice of 'neriage'.

Often, students do not consider the most powerful problem solving approaches without further prompting. We therefore provide students with some "sample student work", chosen and collected by ourselves. This work is designed to show more sophisticated attempts at the problem. Students' task is now to critique, improve, complete and extend suggested solutions – a challenge to their existing thinking.

The concept-focused lessons are similar in structure to the problem solving lessons, but here we identify the different *task genres* that promote concept development and select a rich of that kind. Examples are given in the table below:

The 5 Task Types

Task genres	**Description of tasks**
Classifying and defining	Students devise classifications for mathematical objects, and/or apply classifications devised by others. They discriminate, recognise properties and develop mathematical language and definitions.
Interpreting and translating between multiple representations	Students match cards that show different representations of mathematical objects - words, diagrams, algebraic symbols, tables, graphs. They share interpretations, compare and group the cards in ways that made connections between underlying concepts. The discussion of common 'misconceptions' is encouraged by the inclusion of distracters.
Testing and evaluating mathematical statements and conjectures	Students are given short mathematical statements or generalisations, are asked to make posters that describe their domain of validity and provide examples, counterexamples and explanations to support their decisions.
Creating and solving variants of	Students devise new or variants of existing problems, prepare solutions then challenge other students to solve them. They offer support when the solver

mathematical problems	becomes stuck. This promotes awareness of the structures underlying problems, and focuses attention on the doing and undoing processes in mathematics.
Analysing reasoning and solutions	Students compare different methods for doing a problem, organise solutions and/ or diagnose the causes of errors in solutions. They begin to recognise that there are alternative pathways through a problem, and develop their own chains of reasoning.

These activities are conducted in a collaborative atmosphere, with the teacher acting as a provoker, using the prepared questions to prompt students to argue and refine their interpretations, with a whole class 'neriage' discussion as wrap-up.

Challenges for research and development

Currently, the tasks presented by high stakes examinations and textbooks, (which in the UK are often written by examiners who focus on repetitive practice of examination-type questions) largely determine the types of task that are used within classrooms. We need to challenge this state of affairs at policy level using such classification schemes as we have described above in curriculum documents to describe learning objectives. A vital component, often missing form such documents, is the vivid exemplification that is necessary to show exactly what such tasks might look like.

At a deeper level, further refinement and illustration of the task-types we have described here is needed; in particular, further classroom evidence of their individual impact on teacher and student practices and performances is required.

In addition, research is needed to show how student performances on conceptual and problem solving tasks might be reliably measured and reported. Otherwise examiners and teachers will continue to assess fragments rather than complete performances.

Background and context

While all members of the team contribute to the various aspects of the Shell Centre's work, the authors have played central roles. Malcolm Swan has led the design of tasks and the elicitation of design principles (see e.g Swan and Burkhardt 2012) while Hugh Burkhardt has played a leading role in the development of the analytic frameworks for describing and balancing tasks and the strategic design of tests and their curriculum support (see e.g. Burkhardt 2009). The work has had ongoing national and international support over the past 25 years.

References

Balanced Assessment (1997-99) *Balanced Assessment for the Mathematics Curriculum*, Parsippany, NJ: Pearson Learning, Dale Seymour Publications, 1998-99

Black, P. J., and Wiliam, D. (1998) Assessment and classroom learning *Assessment in Education* 5, 7-74

Burkhardt, H. (1990). *Specifying a national curriculum*. In I. Wirszup, & R.Streit (Eds.), Developments in school mathematics around the world **2** 98-11. Reston, VA: NCTM

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

Cockcroft Report (1982) Mathematics Counts, Report to the UK Government of the committee of enquiry. London, UK: HMSO.

Daro, P. and Burkhardt, H. (2012) *A population of assessment tasks,* J. Mathematics Education at Teachers College 3 Spring-Summer 2012

OFSTED (2012) *Mathematics: made to measure* May 2012, No. 110159, Retrieved from www.ofsted.gov.uk/resources/110159.

Swan, M., (2006) Collaborative Learning in Mathematics: A Challenge to Our Beliefs and Practices, Leicester, UK: NIACE

Swan, M., and the Mathematics Assessment Project team, (2011) materials retrieved from http://map.mathshell.org.uk/materials/lessons.php

Swan, M. and Burkhardt, H. (2012) A Designer Speaks. *Educational Designer*, 2(5). Retrieved from: http://www.educationaldesigner.org/ed/volume2/issue5/article19