Draft – comments welcome

Domain Frameworks in Mathematics and Problem Solving

Alan Bell and Hugh Burkhardt

Shell Centre, Schools of Education
University of Nottingham UK and Michigan State University
with the MARS team

The examples of assessment tasks and student responses were mostly developed by the MARS team for the UK Qualifications and Curriculum Authority as part of the World Class Arena. that illustrate this paper are in Appendix 1 (bound separately in the written version) and linked to references in the text (in the website version, see http://www.nottingham.ac.uk/education/MARS/eval/)

Introduction

In many fields there is an essential complementarity between the analytic and the holistic – for example in music, between the rules of melody and harmony and musical compositions. In assessment, the holistic aspect is represented by the assessment tasks themselves, which provide students with "the opportunity to show what they know, understand and can do"; the complementary analytic framework is provided by the specification of the domain of performance to be assessed. While MARS sees the richness of the task set as the key factor in the quality of an assessment, the domain framework is essential for the explanation of the assessment, and for the balancing of the tests. The challenges in designing such a framework are substantial, when you move beyond short technical exercises assessing knowledge and skills to the assessment of substantial reasoning involving higher level skills.

This presentation will describe the framework for balanced assessment that has been developed by MARS for "Problem Solving in Mathematics, Science and Technology", folowing a decade of work on similar frameworks for Mathematics It will make brief comparisons with that developed by the PISA group on Problem Solving. Each analytic aspect will be illustrated by task exemplars.

These frameworks are multi-dimensional, embracing process, contextual and performance aspects as well as subject content. They are not constrained by traditions in assessment but aim to look for the broader range of performance in the domain which has direct value in life and work outside the classroom. For example, the important novel dimension of Task Type in Mathematics includes Design or Plan, Evaluate, Optimize, Select as well as Find Relations and Deduce from Data.

PISA and MARS

The PISA project and MARS do not differ significantly in their basic definition of problem solving.

Problem solving is cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver. (PISA)

Problem solving is the activity called into play when there is a demand to apply knowledge, skill and experience to unfamiliar situations. (MARS)

The key characteristics of the PISA problem solving assessment are quoted as:

- Focus on ***analytic reasoning*** (including inductive, deductive and critical thinking strategies), which is at the core of problem solving competence and which is the competency that allows students to apply their knowledge and strategies from particular domains across a wide range of exercises;

- Gather information on students' ***processes and strategies*** in solving problems;

- Be framed as a ***cross-curricular competency***, in that assessment tasks will call for the integration of knowledge and strategies from at least two, different curricular areas and not be readily identified by students as related strictly to one discipline;

- Be composed of ***complex tasks***, which i) have multiple variables; ii) are open-ended, as well as goal-directed; and iii) utilise multiple representational modes and cognitive tools;

- Be embedded in ***rich, real-life settings*** that go beyond those normally encountered in school, which will bring the problem solver into novel settings in which he or she is required to transfer knowledge and strategies from one setting to another (and from one subject area to another, as above).

Following this specification, the PISA framework for the assessment of problem solving in a cross-curricular setting focuses on students' competencies in **reasoning critically in both goal-directed and open-ended problem situations.** It is suggested that one project, or theme based assessment block, could include a number of items, possibly of different formats, testing this. Some might be closed, some open items; some might evoke inductive reasoning (for example, trip-planning or provision stocking situations), others deductive reasoning (interpretation of zoning requirements, structuring of proofs, restructuring of common language into more formal arguments). Critical/complex reasoning might be promoted by situations concerning issues of health, risk, environment, and cost-benefit, of population. Complex problem solving could be evoked by real life problems in contexts such as systems analysis or design problems, trouble-shooting problems, or decision-making problems. Some examples of PISA themes and problems are given in Appendix 2

Thus both projects emphasise the testing of complex processes through substantial tasks, but differ somewhat in their approaches. The MARS approach is *primarily holistic*: first we design a collection of tasks which will interest and challenge the target students, then analyse each task in terms of an analytical framework to check the balance of tasks chosen for a particular test. The framework itself is developed by inference from the analysis of worthwhile tasks. The rationale for this approach is the belief that it is crucial to the quality of the assessment that the framework be seen as an illuminative and supportive device, not the prime focus of assessment design. This is because in rich subject areas like Problem Solving (or Mathematics or Science or Technology or .....)

- all analytical descriptions are *weak models of performance* with little predictive or design power;
- in tackling substantial tasks, the strategic load of deciding on, carrying through and monitoring the solution process is demanding, so that the whole is much more than the sum of the parts; the ability to do each step successfully, when asked, does not mean that one can do the whole thing;
- even the separate ingredients are not well-defined, in nature or in difficulty (serious attempts at criterion referencing produce lists of thousands of skills; the challenge of elements in short lists depends strongly on the particular task in which it appears);
- the various elements (strategies and tactics, concepts and skills) interact in determining the difficulty of a task.

Just as a good piece of music is much more than the rules of melody, harmony and rhythm, so a good assessment task asks the student to consider, choose, deploy and integrate various appropriate elements (only crudely described in the domain framework) in a way that forwards the solution.

The PISA approach seems closer to the design process which dominates so much assessment design: first specify the elements of the framework to be tested, then to construct items to meet the specification.

MARS has found it best to encourage designers to design and develop good rich tasks which, viewed holistically by those with well-developed professional judgment in the field, represent opportunities for quality performance in the subject. (Good designers have, of course, already internalized and integrated the essential elements of performance across the domain) The tasks can then be analyzed using the domain framework, and the analysis used to help balance the test, again challenging and/or confirming a holistic initial attempt. (We do not exclude a modest amount of framework driven gap filling!)

In the remainder of this paper we shall describe and illustrate the MARS framework. But it may be appropriate at this point to indicate the differences from Pisa which we expect to emerge.

There are certain differences in purpose and context. The MARS brief is to assess 'Problem-solving in Mathematics, Science and Technology', for high-achieving students of 9 and 13; PISA is intended for 15 year-olds. and the context is cross-curricular and real-life. The styles of task are somewhat different. The PISA tasks appear to be embedded, in the main, in quite sophisticated aspects of adult commercial and public life, whereas MARS has sought to use contexts close to the worlds of the students' outer life, and the inner life of their interests and fantasies. We suggest that, if there exists an ability for problem solving – which we and PISA agree will be somewhat dependent on domain- and context-familiarity - a valid assessment of it needs to call on these more student-friendly contexts.

The MARS framework

The framework which has emerged from the considerations above, and the set of tasks of which these are examples, contains several dimensions. The tasks to be used in explaining this framework, and in the descriptions of student performance that follow, are contained in Appendix 1. The reader is advised to study at least some of these before reading further.

Task Type

The most important heading, task type, characterizes in a holistic way some of the most common occurrences of problem solving in the real world. It attempts to capture the thrust or purpose of the task as a whole, the reason why someone might want to perform it.

Design or Plan

The design of an object, mechanism or structure, or the plan of an activity, involves a creative synthesis of a number of desiderata and constraints; these may include dependence on scientific or quantitative feasibility, or the properties of materials. Often aesthetic judgments are relevant. Often, too, the first stages include the making of a representation of the system, the evaluation of some prototypes or some previous rejected example, the identification of relationships within the system. In the present set of tasks, *Snakes and Ladders* is a typical example, where after reviewing some given faulty game boards, the student has to design a new good one, free of faults. Another is *Pop-Up card,* in which, after some preliminary questions to establish the notation, the student is asked to draw a design for a card to a given specification. Another type of design activity occurs in *Paper Aeroplane* (not shown), where the task is to design an experiment to find the relation between wing span and time of flight.

Evaluate, Optimize, Select

This type differs from Design and Plan in that the outcome is a decision – or rather a supported recommendation for a decision – instead of an object or plan. similar syntheses of desiderata and constraints are needed, but the outcome – and often the criteria, are more one-dimensional. Examples are *Motorway Emergency* (not shown), in which data on speeds and delay times for helicopter and road ambulance have to be analyzed to find the best choice for sending to incidents at different distances; *Holidays,* in which the student has to function as a travel agent, searching a database to find a holiday which fits the requirements of a client; and *Making Soft Toys* (not shown), in which data of cost and time for making two types of toy are given, and the task is to find what number of each should be made to maximize profit.

Model and Estimate or Deduce

In these tasks, the student has to make and/or work with representations of real situations. In *Newspaper,* a double sheet taken from a newspaper is shown, with its page numbers, and the request is to find how many numbered pages are in the whole paper. *Moon Gazing* is an example, where, given a sun-moon-earth picture and pictures of eight moon phases, the two representations have to be matched. As well as working with these, the student has to imagine the motion of the actual objects and relate this to the pictures. In *Bicycle*, the student must make a mental model of the way the cogs and chain work, and draw conclusions.

Review & critique
These are tasks in which a proposed argument or solution is offered for critical comment and suggestions for remedy or improvement. *Skeeters* and *No Smoking*. are examples. They both take the form of checking the consistency of stated hypotheses with some given data – an important part of scientific enquiry that perhaps deserves more explicit recognition.

Deduce from data, fit constraints
This is the classic problem type. Some data are given, and the task is to extract some specified further information by complex inference. We have many examples. *Feeling Unwell* is of this type: several children are pictured, saying what they have eaten; all but one of them have been feeling unwell. The task is to identify the suspect food. Another example is *Eggs* (not shown), in which some rules for giving eggs of different colors to boys & girls in a family are given, and from information on the total number of eggs of each color given next door, the composition of that family is to be inferred. *Triplets* presents three boys, each making a statement such as 'The one in the middle is Tom'. Candidates are required to state which boy is which, and justify the answer. *Voting Results* gives some facts about a class vote for the preferred book (out of three). (34 votes in all, winning book got less than half the votes, a tie for second place). Students have to find all the possible ways in which the votes could have been cast.

Find Relations
This is also almost a standard type of problem or investigation. In mathematics, it is the exploration of a situation to find and prove generalizations, in science it is seeking and establishing principles or theories, either by experiment or by inference from other known principles. In technology it is finding rules for how a given system works. It is a component of many of the more complete types of task above, so is assigned only if this is the sole or dominant objective of the task. In the set described, *Number Pyramid* concerns the mathematical relations embedded in a number pattern, whereas *Pollen* gives data of temperature, humidity and pollen count on different dates, and asks whether and in what way the pollen count appears to depend on the other factors.

Translate – Interpret & Re-present
Here data presented in one symbolic, graphical or diagrammatic form has to be interpreted and (normally) presented in a different form. An example is *Hike,* where, although the explicit task is to find a minimum distance through a road network, the first part the task is the production of a road diagram, starting from a standard

triangular table of distances; there is then a request to find the shortest route from one village, visiting all the rest and returning to the start. *Strange Rock* shows a 3 times table in an 'ancient' notation using geometric symbols for units and groups of 4 and 4^2, and asks students to identify the table and to make translations from and into the 'ancient' symbolism. *Rope* shows a point graph of weight against length for seven pieces of rope, and asks which points represent ropes of the same length, the same weight, and, finally, thick and thin rope; these last require the co-ordination of length & weight.

The other dimensions of the framework are more analytic.

Conceptual Content

Since the brief for our current project specifies Problem Solving in Mathematics, Science and Technology, this heading indicates the balance of the tasks demand among these conceptual fields: *Mathematics, Science* and *Technology.*

Context type
We also wish to specify the provenance of the **context** in which a given task is embedded, since we wish to ensure the maximum relevance of the activity to the experience of the students. So we specify the context type:

student life, adult life, curriculum or *no external context*,

and allocate the task to the first applicable category. Here *student life* means a context with which the student is likely to be somewhat familiar in his life outside the classroom; *curriculum* implies that the context is an element of the school curriculum, which the student has almost certainly met.

Practicality
Under the Context heading we take no account of the familiarity of the task itself, only of the context. The relevance of the question comes into play in the next heading, which describes the **goal type** or **purpose**, namely:

Here we aim to note the extent to which the method and result of the task are *immediately useful*, as distinct from providing *insights* or *methods*, which may become useful in the future. (Many tasks lie between the extremes of this continuum, so we try to indicate their place by partitioning 10 points.)

Open-ness
The degree of open-ness of a task is a relevant feature, and one in which it is hard to reconcile realism with the demands of an assessment. In real life, many problems consist of no more than unease about a situation, or a desire to clarify possibilities or to move towards a decision. In such cases, the first need is to formulate the problem in a way, which lends itself to progress towards a solution. Similarly, the 'solution' may consist of identifying a number of possibilities, maybe with different degrees of desirability. These vital qualities of problem-posing and evaluating solutions are difficult to capture in assessment tasks, on account both of shortness of time and of the need for reliable marking. On the other hand, all the tasks in a problem-solving test are open 'in the middle', in the sense that they all require a non-routine search process to find a solution path, and possibly even a considerable amount of question-formulating and checking in the course of the work. So under this heading we distinguish:

open-middle and *open-ended*

tasks, and also those for which

multiple solutions

are demanded. (We do not count tasks in which there is more than one possible solution, but only one is asked for. The really challenging mental act is, having found one solution, to 'get out of the rut' you have made sufficiently to see other possibilities.)

Reasoning Length
Associated with open-ness, we note the *reasoning length*, which is the estimated time devoted to the longest unprompted part of the task. This distinguishes a long task consisting of a number of separate short tasks from one, which demands sustained

exploration or reasoning. Since the ability to creat and sustain substantial chains of reasoning is an important aspect of performance in problem solving (and many other fields) this is a key variable.

Phases
Within the problem-solving process it is common to identify certain phases. The classic description of the applied mathematical process consists of making a mathematical model of the situation, manipulating within the model, interpreting the solution in terms of the original situation, evaluating its plausibility, then refining and repeating the cycle as necessary. There is also Polya's somewhat similar four-part scheme of understanding the problem, making a plan, executing the plan and looking back. In the equivalent science process, the middle stages would be replaced by the design and conduct of an experiment (which could be a thought-experiment); in technology, an object or mechanism might be designed and made.
In practice, we find a number of these phases occurring at all stages of the problem-solving activity. For example, it is rare for an overall plan to be articulated at the beginning of a solving activity, but common for small-scale plans to be frequently made and executed. Similarly, the reflective evaluation of a completed solution does not, we believe, commonly happen in student problem-solving; and even if it does, there is rarely evidence of it in the student's script, unless it is specifically demanded, in which case it cannot be credited as a chosen act.
We have to choose what characteristics of the problem-solving process are sufficiently important for us to use them to classify tasks. The aspects we distinguish under this heading are:

Formulating, Processing, Interpreting, Checking, Reporting.

We certainly wish to know what degree of formulation is needed, and whether there is need for significant interpretation of the outcome of the processing. (Even between these two it is difficult to distinguish, since a problem may be so well formulated initially that there may be no need for interpretation as such of the outcome – the outcome gives the answer.) We can also identify and reward the presentation of an argued report of the work and its conclusions, if this is asked for, and this we do.

In addition to these *dimensions of balance*, we sometimes use an additional heading of **non-routine-ness**, to specify whether the task is non-routine in *method-or-inference* (all problem-solving tasks are), and/or in *context*.

What characterizes the problem solving performance of able students?

This is the subject of the sequel to this paper.

Appendix 1

Some MARS tasks and student work

This appendix is bound separately.

Appendix 2

Some PISA tasks + notes

Polluting Factory (Design problem)

This problem is organized around an air pollution issue that requires judging and comparing information from 5 different sources:

1 Geographical Map of a section of a town with a factory (pollution source), a residential area, a shopping area, School A, School B, and trees
2 Newspaper Article
3 Table of values of polluting chemicals
4 TV Interview with the chief operating officer
1 City Council Costs for:
**Building a new school at a different place *Introducing new technology in the factory *Planting trees around the factory *Building a higher chimney *Closing the factory*

Short questions are used to test comprehension, and judgments of the likely truth, of the information given; then more open questions ask for arguments for and against the proposed actions, and for plans for economic use of available funds over four years.

Well Drilling in Rural Areas (Systems analysis problem)

This question gives a graph showing the yearly water supply obtained by a rural community over a 30 year period during which two wells were dug. Short questions test comprehension, then more open ones seek reasons which might be given to explain the events by local people with different perspectives – water engineer, ecology activist, boss of supply station; another asks what further information would be required to decide whether the water supply meets their needs.

Bicycle Pump1 (Trouble shooting Problem)

A diagram of a pump is given, and explanations are requested of how the pump operates, and, if the pump fails to pump air, what might be wrong.
An alternative version gives a verbal description of the pump, and asks for an explanatory diagram.

Buying a Mobile Phone (Decision making Problem)

Advertisements are shown for two phones, and three service plans. The student is asked to suggest advice which might be offered by a shop worker; and to give their own advice on the best buy.