Goodchild, S. (Ed.) Proceedings of the British Society for Research into Learning Mathematics 22(1&2) February/May 2002

USING AND APPLYING MATHEMATICS: DESIRABLE TEACHING STYLE OR ASSESSMENT CONSTRUCT?

Cathy Smith

Homerton College, University of Cambridge

From 1994 to 1997 schools could choose between two GCSE syllabus options for assessing Using and Applying Mathematics: coursework tasks or written Ma1 papers. This paper sets out to examine the effect of this choice on the nature of the assessment. In attempting to establish methods of comparing assessment instruments without reference to candidates' performance, I trace the origins of these modes in 'authentic' and 'competence-based' assessment. Scrutiny of the papers and examiners' reports suggests that the original purposes underlying the choice of mode have been subjugated to the examiners' requirement for reliability, and that the combination of new assessment instruments with traditional examining processes can lead to less valid assessment.

INTRODUCTION

The starting point for this enquiry was my feeling as a then secondary teacher that GCSE Mathematics syllabus options were so different in their assessment of Ma1 that they were not comparable. The project then developed: to research current thinking on assessment in order to find a structure or a rationale on which I could base a more objective comparison. The resulting work is an attempt to make a bridge between thinking about assessment theories and the evaluation of some actual assessment materials.

ASSESSMENT MODES

The syllabus chosen was the London GCSE Mathematics syllabus 1384, a popular syllabus whose options are fairly extreme examples of the typical choice available. There are two main methods for assessing achievement in Using and Applying Mathematics (or what might be called mathematical processes):

	Options A,B,C1: Coursework	Option C2: Written papers
	Portfolio of open ended tasks or 2 timed tasks	2 half hour papers (5 questions on a common theme)
marked by:	Performance indicators	Markscheme

The assessment of Maths content (Ma 2-5) is by common written papers.

An example of the type of (timed) task set is "Dice Roll". Imagine placing a die on a table with the 2 facing upwards. Choose a direction and roll the die over a quarter turn. Do this 2 times - because 2 is the number on top - in the same direction, and the die stops

© BSRLM and Others

with 5 on top. Choose a direction again and roll the die 5 times. Investigate the sequences of stopping numbers.

The written Ma1 papers were tiered, with the same contextual theme, e.g. stopping distances, for all tiers. The following 1997 question is at NC level 6, common to all tiers.

STOPPING DISTANCES

Speed (s)	Braking distance (b)	Stopping distance (d)
20	20	40
30	45	75
40	80	120
50	125	175
60	180	240

Q. Investigate the value of $\frac{b}{s}$ for different speeds in order to predict the stopping distance of a car at 80mph.

THEORETICAL PERSPECTIVES

Comparable standards

Cresswell (1996) offers a rare but comprehensive perspective on the issue of comparable standards in public examinations. Quantitative definitions based on corresponding groups of candidates attaining identical grade distributions, citing the impossibility of equating sensitive variables such as ability, motivation, timing, school background. He argues further that no statistical procedure can allow for a difference in *what* is being assessed, nor recognise that a particular syllabus organisation may lead to genuinely better performance. Cresswell then proposes a social definition based on value judgements formed by awarders (e.g. the GCSE examiners) accepted as competent by all interested users.

Comparing assessment instruments

A comparison similar to that I am attempting is that of Brown (1998) and Brown and Moore (1994) with regard to the different options for assessing practical and investigational skills in A level Biology. Statistical correlation and factor analysis is used to determine whether, for the different options, there is a distinguishable construct that is assessed by the practical/investigational element, and not by the other papers.

© BSRLM and Others

Underlying assessment paradigms

Explicit assessment of mathematical processes in GCSE was a new requirement of the 1991 National Curriculum which identified Using and Applying Mathematics as 20% of the curriculum. One purpose was that of system control - to ensure that this part of the curriculum was actually taught. At the same time, there is a clear influence from two then developing bodies of thought recommending the validity of *authentic* or *competence based* assessment. In both of these schools of thought assessment may use context-related tasks, and has an referent external to the immediate aims of schooling.

PRINCIPLES

The theoretical perspectives above suggested three principles from which to structure my comparison. Firstly it led me to look further than statistics on performance. In 1997 not only the grade distributions but the grade boundaries for the different syllabus options are very similar, leading to an apparent dead end. However, these are only the end results of the process of making judgements. Awarders communicate and justify their decisions about assessment in Examiners' Reports and so these are equally a valid source of evidence for analysis.

Secondly a difference can lie not only in *how* these options assess Ma1 but in *what* is being assessed. The Examiners' Report and the materials themselves can be examined for a direct comparison. As in the A level Biology studies an approach is to investigate the relationship of each Ma1 assessment to the assessment of content in the Ma2-5 papers.

A third relevant comparison is the extent to which each assessment instrument fulfils its intended purpose. Again, the Examiners' Reports will serve to illustrate whether in practice the goals of either competence based or authentic assessment are achieved.

COMPARISON

Criterion referencing

Central to competence based assessment is the two-fold aim of describing all aspects of desired performance, and of making these performance criteria as transparent and unambiguous as possible (Wolf,1996). The 1991 National Curriculum with its attainment targets, strands and individual statements of attainment (SoAs), and equating of curriculum with assessment, is an obvious example of criterion referencing, and the GCSE syllabus follows this structure for both mathematics content and processes. There is already much debate about the possbility of *valid* assessment of higher-order mathematical skills with the use of a set of isolated performance descriptions (Wolf 1996, Hewitt,1992). GCSE examiners argue further that context-free descriptions of performance, such as SoAs, are not *reliable* and so insist on the use of criteria written

© BSRLM and Others

specifically for the task set (Cresswell and Houston,1991).

For the GCSE coursework option competence-based assessment is carried out by this accepted practice of publishing performance indicators for each coursework task. For example, a Dice Roll performance indicator for SoA 6a "Pose their own questions or design a task in a given context" is

> The candidate begins to generate sequences of STOP numbers by going beyond one sequence for each of the starting positions (except 4). There should be evidence of the candidate trying to work systematically.

However the evidence from the Examiners' Report suggests that the perfomaance indicators are in fact a distraction. A main cause of unsatisfactory teacher assessment is "too literal use of the Performance Indicators as a mark scheme" (p65). Equally there is no mention of individual SoAs being used to form judgements. The nearest the Examiners' Report comes to discussing agreed criteria is in its mention of NC strands but again only 7 out of 23 paragraphs refer even to these, with some of these stressing the importance of "common sense" judgements e.g. that 'reasoning' is unlikely to score higher than 'doing'. Most of the comments in the report are expressed in terms of what the more /less able candidate might achieve. It is clear that in practice the assessment is based not on criteria but on the examiners' holistic judgement of standards conveyed implicitly through the performance indicators and the moderation procedures. Successful teacher assessment occurs when the teachers share these standards.

In the written paper Ma1 option questions are matched to individual SoAs. These are "randomly sampled" from the curriculum to permit comprehensive coverage while allowing some coherence to the themed paper. Assessment is by markscheme eg

SoA	Working	Marks	Notes
6a,	Sequence of 1,1½, 2, 2½, 3	b2	any 2/3 correct is b1
b,c	BD(80) =4x80	m1	sight of 320 is b3
	SD(80)=320+80=400	a1	400 identified is b4

Although for individual questions such a markscheme is unambiguous, the use of a grading system based on aggregated marks in itself runs counter to the goal of providing assessment information about specific criteria. All questions are equally weighted whatever the level of the corresponding SoA while sampling of SoAs is in practice constrained by tiering and the "common sense" principle - questions on 'doing' usually come before and so are at a lower level than questions on 'reasoning'. Moreover the same response to almost identically phrased questions in the same paper can get different marks depending on which SoA is being assessed (e.g. numerical testing in Paper 20 Q2b and 5).

© BSRLM and Others

Neither of the assessment options fulfills the purposes of competence based assessment. The coursework option is seen to be marked by value judgements rather than criteria, while the written paper uses criteria only in the paper design and in a manner that distorts the intended curriculum.

A separate assessment construct?

The second comparison sets out to establish whether in each syllabus option Using and Applying Mathematics is being assessed as a separate construct from mathematical content.

The look and vocabulary of the Ma 1 written papers is clearly similar to that of the content papers but there is explicit direction in the content questions (also matched to SoAs) on which methods to use. The markschemes however show that the responses expected in the different papers are very similar. Indeed there are several examples of the same marks being given for the same working on content and processes papers (eg Paper6 Q9 and Paper 21 Q4). Thus the only difference in what is being assessed lies in the candidates' ability to choose a strategy. In practice the questions set involve only a small range of techniquesand the examiners report that candidates rely on an even smaller selection in their answers.

The coursework tasks do not flag their links to mathematical content but the examiners clearly have a prior knowledge base in mind. For example not knowing that a square is a rectangle bars achievement above level 5, and sight of an algebraic variable gives level 7. Use of content is not however necessary, for instance Foundation and Intermediate candidates were found to score unusually highly in Dice Roll Strand 3 precisely because they could explain cycles in relation to a physical situation without using algebra (Examiners' Report p67). At higher levels the examiners' reliance on mathematical content to communicate standards is increasingly clear. The report on Dice Roll describes achievements at higher levels largely in terms of using modular arithmetic and route matrices, neither of which are even in the content syllabus (hence the very few higher level achievements).

There is little evidence from the Examiners' Report that either syllabus option assesses mathematical processes in isolation from content, except for the lower grades of coursework task. The written papers are perhaps more honest than the coursework in displaying the link with content papers but not more successful. Few candidates are able to tackle the hardest problems in either option.

Authentic achievement

Authentic academic achievement is first defined by Newmann and Archbald (1992) as production instead of reproduction of knowledge, disciplined enquiry dependent on a prior knowledge base, and having value beyond assessment. The inseparability of problem solving and content knowledge base is emphasised in this original definition

© BSRLM and Others

although not in other interpretations of authentic assessment.

It is here that the clearest difference between the written paper and coursework tasks may be seen. The Ma1 questions primarily require candidates to recognise standard tasks. They are similar to traditional exam questions, and contrast with the content questions only because in the 1997 papers these are fragmented and directive.

The coursework tasks at their best allow productive enquiry in nonstandard complex problems which may be more realistic to demands beyonds schooling. This work confirmed the extent to which assessment relies on examiners judgements rather than curriculum descriptions. The nature of the coursework option is then best seen by whether the average performance expcted by the examiners is an authentic academic achievement. The examiners' repeated advice at these levels that candidates learn set techniques, eg tabulation, puts in doubt both the productive nature of the work and its value beyond the assessment.

REFERENCES

Brown, C. (1998) 'An evaluation of 2 different methods of assessing independent investigations in an operational pre-university level examination in biology in England' *Studies in Educational Evaluation*, 24(1).

Brown, C. & Moore, J. (1994) 'Construct validity and context dependency of the assessment of practical skills in an advanced level biology examination' *Research in Science and Technological Education*, 12(1): 53-62.

Cresswell, M. (1996) 'Defining, setting and maintaining standards in curriculum-embedded examinations; judgements and statistical approaches' in Goldstein, H. & Lewis, T. (Eds*) Assessment: Problems, developments and statistical issues.* Chichester: Wiley, pp. 57-84.

Cresswell, M. & Houston, J. (1991) 'Assessment of the National Curriculum - some fundamental considerations' *Educational Review*, 43(1): 63-78.

Hewitt, D. (1992) 'Trainspotter's paradise' *Mathematics Teaching*, 140.

Newmann, F. & Archbald, D. (1992) 'The nature of authentic academic achievement' in Berlak, H., Newmann, F., Adams, E., Archbald, D., Burgess, T., Raven, J. & Romberg, T. (Eds*) Towards a new science of educational testing and assessment.* Albany NY: State University of Albany Press, .

ULEAC (1997) *Syllabus: GCSE Mathematics, Subject report and mark schemes.* Edexcel Foundation.

Wolf, A. (1996) 'Vocational assessment' in Goldstein, H. & Lewis, T. (Eds*) Assessment: problems, developments and statistical issues.* Chichester: Wiley, pp. 209-230.

© BSRLM and Others