

Formative Assessment in Mathematics

Part 1: Rich questioning

Dylan Wiliam

Equals: Mathematics and Special Educational Needs 5(2) 15-18; Spring 1999

Introduction

Successive governments have bemoaned the 'long tail of underachievement' in British schools, and the clear implication of such a phrase is that achievement in Britain is skewed towards the lower end. In fact, the distribution of achievement in British schools is almost completely symmetrical, and what skew there is towards the higher end (would we call that a long tail of overachievement?). While there is no evidence of a skewed distribution, however, it is true that the range of achievement in Britain is wider than in almost any other developed country. Our highest-performing students compare well to the best in any other country, but we have many students who leave school or college without adequate capability in mathematics.

Now the typical argument made by politicians is that this is unacceptable because the lack of an adequately skilled workforce harms our industrial competitiveness, but this argument simply doesn't hold up, because as many studies have shown, there is no discernible association between levels of academic achievement and industrial productivity.

Nevertheless, I believe we should be concerned about the levels of mathematical achievement of school leavers in this country. The reason for this is that in my view too many of our young people leave school without the mathematical capabilities they need in order to exercise an acceptable degree of control over their own lives.

About a year ago, Paul Black and I published a review of approximately 250 studies, carried out over the last ten years, into the effectiveness of formative assessment in raising standards of achievement [1]. What we found was that increasing the use of formative assessment in school classrooms does produce significant increases in students' learning—enough to raise levels of performance in mathematics amongst British students to fifth place in the international 'league tables' of mathematical performance, behind only Japan, Singapore, Taiwan and South Korea. Put another way, appropriate use of formative assessment would raise the average achievement students by as much as 2 grades at GCSE.

But much more importantly, formative assessment has the power to change the *distribution* of attainment. Good formative assessment appears to be disproportionately beneficial for lower attainers, so that typically, an average improvement of two GCSE grades would actually be an improvement of three grades for the weakest students, versus an improvement of one grade for the strongest. Formative assessment therefore seems to be the most promising way to reduce the unacceptably wide variation in attainment that currently exist in mathematics classrooms in Britain.

Parts 2 and 3 of this article will appear in future issues of 'Equals' and will deal with giving feedback to learners, the importance of sharing learning goals with students, and student self-assessment. The main focus of this first part is with the use of questions to support learning.

What makes a good question?

Two items used in the Third International Mathematics and Science Study (TIMSS) are shown in figure 1 below. Although apparently quite similar, the success rates on the two items were very different. For example, in Israel, 88% of the students answered the first items correctly, while only 46% answered the second correctly, with 39% choosing response (b). The reason for this is that many students, in learning about fractions, develop the naive conception that the largest fraction is the one with the smallest denominator, and the smallest fraction is the one with the largest denominator. This approach leads to the correct answer for the first item,

but leads to an incorrect response to the second. In this sense, the first item is a much weaker item than the second, because many students can get it right for the wrong reasons.

Item1 (success rate 88%)

Which fraction is the smallest?

- a) $\frac{1}{6}$ b) $\frac{2}{3}$ c) $\frac{1}{3}$ d) $\frac{1}{2}$

Item 2 (success rate 46%)

Which fraction is the largest?

- a) $\frac{4}{5}$ b) $\frac{3}{4}$ c) $\frac{5}{8}$ d) $\frac{7}{10}$

Figure 1: two items from the Third International Mathematics and Science Study

This illustrates a very general principle in teachers' classroom questioning. By asking questions of students, teachers try to establish whether students have understood what they are meant to be learning, and if students answer the questions correctly, it is tempting to assume that the students' conceptions *match* those of the teacher. However, all that has really been established is that the students' conceptions *fit*, within the limitations of the questions. Unless the questions used are very rich, there will be a number of students who manage to give all the right responses, while having very different conceptions from those intended.

A particularly stark example of this is the following pair of simultaneous equations:

$$3a = 24$$

$$a + b = 16$$

Many students find this difficult, often saying that it can't be done. The teacher might conclude that they need some more help with equations of this sort, but the most likely reason for the difficulties with this item is not to with mathematical skills but with their *beliefs*. If the students are encouraged to talk about their difficulty, they often say things like, "I keep on getting b is 8, but it can't be because a is". The reason that many students have developed such a belief is, of course, that before they were introduced to solving equations, they were almost certainly practising substitution of numbers into algebraic formulas, where each letter stood for a different number. Although the students will not have been taught that each letter must stand for a different number, they have generalised implicit rules from their previous experience (just as because we always show them triangles where the lowest side is horizontal, they talk of "upside-down triangles").

The important point here is that we would not have known about these unintended conceptions if the second equation had been $a + b = 17$ instead of $a + b = 16$. Items that reveal unintended conceptions—in other words that provide a "window into thinking"—are difficult to generate, but they are crucially important if we are to improve the quality of students mathematical learning.

Now some people have argued that these unintended conceptions are the result of poor teaching. If only the teacher had phrased their explanation more carefully, had ensured that no unintended features were learnt alongside the intended features, then these misconceptions would not arise.

But this argument fails to acknowledge two important points. The first is that this kind of over-generalisation is a fundamental feature of human thinking. When young children say things like "I goed to the shop yesterday", they are demonstrating a remarkable feat of

generalisation. From the huge messiness of the language that they hear around them, they have learnt that to create the past tense of a verb, one adds “ed”. In the same way, if one asks young children what causes the wind, the most common answer is “trees”. They have not been taught this, but have observed that trees are swaying when the wind is blowing and (like many politicians) have inferred a causation from a correlation.

The second point is that even if we wanted to, we are unable to control the student’s environment to the extent necessary for unintended conceptions not to arise. For example, it is well known that many students believe that the result of multiplying 2.3 by 10 is 2.30. It is highly unlikely that they have been taught this. Rather this belief arises as a result of observing regularities in what they see around them. The result of multiplying whole-numbers by 10 is just to add a zero, so why shouldn’t that work for all numbers? The only way to prevent students from acquiring this ‘misconception’ would be to introduce decimals before one introduces multiplying single-digit numbers by 10, which is clearly absurd. The important point is that we must acknowledge that what students learn is not necessarily what the teacher intended, and it is essential that teachers explore students’ thinking before assuming that students have ‘understood’ something.

Now questions that give us this “window into thinking” are hard to find, but within any school there will be good selection of rich questions in use—the trouble is that each teacher will have her or his stock of good questions, but these questions don’t get shared within the school, and are certainly not seen as central to good teaching.

In Britain, most teachers spend most of their lesson preparation time in marking books, invariably doing so alone. In some other countries, the majority of lesson preparation time is spent planning how new topics can be introduced, which contexts and examples will be used, and so on. This is sometimes done individually or with groups of teachers working together. In Japan, however, teachers spend a substantial proportion of their lesson preparation time working together to devise questions to use in order to find out whether their teaching has been successful.

Now in thinking up good questions, it is important not to allow the traditional concerns of reliability and validity to determine what makes a good question. For example, many teachers think that the following question, taken from the Chelsea Diagnostic Test for Algebra, is ‘unfair’:

Simplify (if possible): $2a + 3b$

This item is felt to be unfair because students ‘know’ that in answering test questions, you have to do some work, so it must be possible to simplify this expression, otherwise the teacher wouldn’t have asked the question. And I would agree that to use this item in a test or an examination where the goal is to determine a student’s achievement would probably not be a good idea. But to find out whether students understand algebra, it is a very good item indeed. If in the context of classroom work, rather than a formal test or exam, a student can be tempted to ‘simplify’ $2a + 3b$ then I want to know that, because it means that I haven’t managed to develop in the student a real sense of what algebra is about.

Similar issues are raised by asking students which of the following two fractions is the larger:

$$\frac{3}{7} \qquad \frac{3}{11}$$

Now in some senses this is a ‘trick question’. There is no doubt that this is a very hard item, with typically only around one 14-year old in six able to give the correct answer (compared with around three-quarters of 14-year-olds being able to select correctly the larger of two ‘ordinary’ fractions). It may not, therefore, be a very good item to used in a test of students’ achievement. But as a teacher, I think it is very important for me to know if my students think that $\frac{3}{11}$ is larger than $\frac{3}{7}$. The fact that this item is seen as a ‘trick question’ shows how deeply ingrained into our practice the summative function of assessment is.

A third example, that caused considerable disquiet amongst teachers when it was used in a national test, is based on the following item, again taken from one of the Chelsea Diagnostic Tests:

Which of the following statements is true:

- (1) AB is longer than CD
- (2) AB is shorter than CD
- (3) AB and CD are the same length



Again, viewed in terms of formal tests and examinations, then this may be an unfair item, but in terms of a teacher's need to establish secure foundations for future learning, I would argue that this is entirely appropriate.

Rich questioning, of the kind described above, provides teachers not just with evidence about what their students can do, but also what the teacher needs to do next, in order to broaden or deepen understanding.

Classroom questioning

There is also a substantial body of evidence about the most effective ways to use classroom questions. In many schools in this country, teachers tend to use questions as a way of directing the attention of the class, and keeping students 'on task', by scattering questions all around the classroom. This probably does keep the majority of students 'on their toes' but makes only a limited contribution to supporting learning. What is far less frequent in this country is to see a teacher, in a whole-class lesson, have an extended exchange with a single student, involving a second, third, fourth or even fifth follow-up question to the student's initial answer, but with such questions, the level of classroom dialogue can be built up to quite a sophisticated level, with consequent positive effects on learning. Of course, changing one's questioning style is very difficult where students are used to a particular set of practices (and may even regard asking supplementary questions as 'unfair'). And it may even be that other students see extended exchanges between the teacher and another student as a chance to relax and go 'off task', but as soon as students understand that the teacher may well be asking them what they have learned from a particular exchange between another student and the teacher, their concentration is likely to be quite high!

How much time a teacher allows a student to respond before evaluating the response is also important. It is well known that teachers do not allow students much time to answer questions, and, if they don't receive a response quickly, they will 'help' the student by providing a clue or weakening the question in some way, or even moving on to another student. However, what is not widely appreciated is that the amount of time between the student providing an answer and the teacher's evaluation of that answer is much more important. Of course, where the question is a simple matter of factual recall, then allowing a student time to reflect and expand upon the answer is unlikely to help much. But where the question requires thought, then increasing the time between the end of the student's answer and the teacher's evaluation from the average 'wait-time' of less than a second to three seconds, produces measurable increases in learning (although increases beyond three seconds have little effect, and may cause lessons to lose pace).

In fact, questions need not always come from the teacher. There is substantial evidence that students' learning is enhanced by getting them to generate their own questions. If instead of

writing an end-of-topic test herself, the teacher asks the students to write a test that tests the work the class has been doing, the teacher can gather useful evidence about what the students think they have been learning, which is often very different from what the teacher thinks the class has been learning. This can be a particularly effective strategy with disaffected older students, who often feel threatened by tests. Asking them to write a test for the topic they have completed, and making clear that the teacher is going to mark the question rather than the answers, can be a hugely liberating experience for many students.

Some researchers have gone even further, and shown that questions can limit classroom discourse, since they tend to demand a simple answer. There is a substantial body of evidence the classroom learning is enhanced considerably by shifting from asking questions to making statements. For example, instead of asking “Are all squares rectangles”, which seems to require a ‘simple’ yes/no answer, the level of classroom discourse (and student learning) is improved considerably by framing the same question as a statement—“All squares are rectangles”, and asking students to discuss this in small groups before presenting a reasoned conclusion to the class.

Conclusion

Over thirty years ago, David Ausubel argued that the most important factor influencing learning is what the learner already knows, and that the job of the teacher was to ascertain this and to teach accordingly. Since then it has become abundantly clear that students’ naive conceptions are not random aberrations, but the result of sophisticated and creative attempts by students to make sense of their experience. Within a normal mathematics classroom, there is clearly not enough time for the teacher to treat each student as an individual, but the good news is that the vast majority of the naive conceptions are quite commonly shared, and as long as the teacher has a small battery of good questions it will be possible to elicit the most significant of these misconceptions. If there teacher does then have any time to spend with individual students, these can be targeted at those whose misconceptions are not commonly shared. After all, teaching is interesting because students are so different, but it is only possible because they are so similar.

References

- 1 Black, P. J. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, 5(1), 7-73.
- 2 Hart, K. M.; Brown, M. L.; Kerslake, D.; Küchemann, D. & Ruddock, G. (1985). *Chelsea diagnostic mathematics tests*. Windsor, UK: NFER-Nelson.

Dylan Wiliam works at King’s College London School of Education. If you have any questions relating to formative assessment that you would like to see discussed in future issues of Equals, please email him at dylan.wiliam@kcl.ac.uk.