Evidence-based Practice in Science Education (EPSE) Research Network

Using Diagnostic Assessment to Enhance Teaching and Learning

A Study of the Impact of Research-informed Teaching Materials on Science Teachers' Practices

Robin Millar and Vicky Hames

September 2003

Acknowledgements

The research project discussed in this report was one of four inter-related projects, exploring the relationship between research and practice in science education, undertaken by the *Evidence-based Practice in Science Education (EPSE)* Research Network, a collaboration involving the Universities of York, Leeds, Southampton and King's College London.

The EPSE Network was funded by the UK Economic and Social Research Council (ESRC) in Phase 1 of the *Teaching and Learning Research Programme (TLRP)* (grant no. L139 25 1003).

We acknowledge the contribution of our EPSE colleagues, John Leach, Jonathan Osborne, Mary Ratcliffe, Hannah Bartholomew, Andy Hind and Jaume Ametller to our thinking about the issues involved in this work.

1 Aim and rationale

The central issue which the project discussed in this report set out to explore is the relationship between research and practice in science education – more specifically, the relationship between research on the learning of scientific ideas on the one hand, and teachers' actions and choices in the classroom on the other.

Over the past 30 years, a great deal of research has been carried out in many countries on learners' ideas about the natural world. This has helped identify commonly-held ideas which differ from the accepted scientific view, and has shown that these are often very resistant to change (see, for example, Driver et al., 1994; Pfundt & Duit, 1994). These findings have clear implications for the pace and sequence of instruction in many science topics, particularly those which involve understanding of fundamental ideas and models. Yet whilst many teachers know of this research, it has had little systematic impact on classroom practices, or on science education policy, in the UK.

The aim of this project was to explore a possible way of enhancing the impact on teachers' practices of the insights and findings of research on science learning.

2 Context and background

Some critiques of educational research have attributed its lack of impact on practice, at least in part, to poor communication between researchers and practitioners (for example, Hillage et al., 1998). In response, some have suggested that researchers need to write up their findings in briefer and more accessible formats, such as short research briefings, which communicate the principal findings of their work clearly and quickly to busy 'users'. Effective communication may, however, involve rather more than this. The steps of identifying the practical implications of research findings, devising materials and approaches to implement them, and testing these in use are far from trivial. Compared to the large body of research on students' ideas in science, rather less research has focused on testing possible ways of improving students' learning of these difficult ideas. Implications identified in the concluding sections of research articles, and the suggestions for action proposed, are usually based on professional experience, rather than on research evidence that the suggested approach will achieve better outcomes than the current one. In fact, as Lijnse (2000) points out, science education research offers little direct or specific guidance to teachers about how best to teach specific topics to specific groups of learners.

There are, of course, some examples of teaching sequences and programmes for school science topics that have been developed by researchers in the light of their findings (and those of others) (for example, CLISP (Children's Learning in Science Project), 1987; Klaassen, 1995; Viennot and Rainson, 1999; Tiberghien, 2000). Viennot (2001: 36-43) discusses an example of research influencing the detail of a national curriculum specification. All of these examples involve a 'transformation' of knowledge – from the form of a research finding (a summary statement grounded in empirical evidence) into teaching materials or guidelines that can be implemented in the classroom.

In this project, we have also chosen the approach of providing teachers with materials that they can use directly. Rather than attempt to communicate research *findings* to teachers, or to develop with them teaching sequences based on research findings, we have produced, and made available to teachers, a collection of *instruments and tools* of the kind used by researchers to collect evidence of students' learning. These diagnostic questions make it easier for teachers to collect data on the progress of their own classes. Their practice is then more 'evidence-informed' in the sense that they can base decisions about the pace and sequence of instruction on better evidence of their students' current ideas and understandings. It is also evidence-informed in the sense that the questions themselves embody insights, and reflect outcomes, of a body of previous research, by focusing attention on key points and issues that have been shown by research to be important for learning.

One influence on the choice of this approach was the evidence of significant impact, on university-level physics teaching in the USA, of the Force Concept Inventory (Hestenes et al., 1990) and several other similar 'inventories' that have appeared in the past decade. These provide a quick means of surveying the understanding of students. Their existence has stimulated many teachers to modify their teaching methods and approaches, in response to what they perceived as unsatisfactory performance by their own students (Mazur, 1997; Redish, 2003) – leading in some cases to measured gains in performance as a result of the changes introduced. One aim of the project reported in this paper is to work towards similar instruments for use at school level.

A second influence was the major research review by Black and Wiliam (1998a, b) showing that the use by teachers of formative assessment can lead to significant learning gains by their students. One barrier to the wider use of such approaches may be the shortage of suitable questions and tasks for formative assessment. If so, providing materials may be significant in increasing uptake of an approach which research has shown to 'work'. This, then, is a third sense in which the teaching methods which this project is aiming to promote are evidence-informed.

3 Overview of the project

In outline, then, the strategy in this project was to develop banks of diagnostic questions that teachers could use when teaching some important science topics, to give these to a sample of teachers, and monitor with them how they used them in their teaching.

The three topics chosen were: electric circuits, forces and motion, and particle models of matter. This choice was made in collaboration with a 'partnership group' of teachers, researchers and other science education practitioners (Local Education Authority science advisors, textbook authors, examiners) who contributed throughout the project to its design and implementation. These topics met the following selection criteria: they are central to the Science National Curriculum for England (DfEE/QCA,1999); they are topics for which it was felt many teachers would welcome additional teaching resources and ideas; they involve understanding models and explanations, rather than recall of facts; there is a substantial body of research on student

learning in these topics, which provides a starting point for developing a diagnostic question bank.

To develop diagnostic question banks, we first reviewed the published literature in these topic areas to collect together as many as possible of the instruments and tools used by researchers to probe pupils' understandings (for example, APU, 1988-9; Shipstone et al., 1988; Hestenes et al., 1990). Where there were gaps, new questions were devised. On the advice of the teachers in the group, all were designed to be relatively quick to use, to facilitate their use in formative assessment, where quick interpretation of a class's responses is essential. For this reason, many of the items in the banks are closed format, such as single-tier and two-tier multiple choice questions (Haslam and Treagust, 1987; Treagust, 1988), though many also have open-response sections. Whilst all could be used as individual written tasks, some are intended as stimuli for group discussion activities or even for predict-observe-explain practical tasks (White and Gunstone, 1992).

Draft questions were first reviewed by teachers in the Partnership Group (and some others) to check their face validity, to improve them where necessary and make them more usable in classrooms. Sets of questions were then piloted in these teachers' schools by class sets of around 30 pupils. We also interviewed a sample of these pupils to explore the reasoning behind their answers. We also trailed some open-response questions to help devise answer options for structured-response versions.

Checks on consistency of response to several items testing the same science idea, or to the same question on two different occasions, were also carried out. This resulted in large banks of diagnostic questions, eventually consisting of over 200 items each for electric circuit and forces and motion, and over 100 on matter and the particle model[1].

Teachers in a sample of 10 schools (8 secondary; 2 primary), were then given the complete item bank on one science topic, with some outline suggestions on possible ways of using these. No training was provided as we did not want to influence the use of the materials too strongly. Four of these teachers had been involved in developing the banks; others had heard about the project and expressed an interest in being involved in it. In most schools, the materials were used by several teachers in the science department. As the project proceeded, several schools also requested a second item bank in addition to the one initially provided.

Over a school year, we then monitored how the teachers chose to use these resources with their classes, their views on them, and any evidence of impact on teaching and/or learning. Teachers were asked to keep records of their use of the materials, and were interviewed about how they were using them, and their impressions. Written questionnaires were used to collect additional information and views. Some teachers also collected samples of pupils' work, or test results. In this way we built up a set of case studies of the use of these diagnostic question banks in 10 schools. In all, 23 teachers provided feedback on their use of the materials, of whom 16 were interviewed in depth (including at least one from every school), several on two occasions during the year. Whilst this sample is relatively small, it included a range of school sizes and types

[1] A sample of these can be downloaded from the EPSE Network web site. The URL is: http://york.ac.uk/depts/educ/projs/EPSE

(city, suburban, small town). The group of teachers (13 male, 10 female) included biology, chemistry and physics specialists, with teaching experience ranging from one year to over 25 years. So, although the sample was not selected randomly, the schools and teachers involved were broadly representative of schools and science teachers in England more generally.

4 The diagnostic question banks in use

In this section we describe and discuss these teachers' use of the diagnostic materials they were given. The account is in five parts. In section 4.1, we report on the different ways in which teachers chose to use the materials. Section 4.2 then discusses teachers' views on the diagnostic materials and their characteristics. The following three sections then look at the impact of the materials on teachers' classroom actions and choices (4.3), on teachers' own understanding of science (4.4) and on pupil learning (4.5).

4.1 Ways of using the diagnostic question banks

4.1.1 End of topic tests

Not surprisingly, many teachers' initial perception of the diagnostic question packs was that these provided the raw materials for making up tests, or modifying existing tests. The initial reaction of one secondary teacher, when given the materials by his Head of Department, was:

> T1[2]: … oh good, questions, because we're so short of useful questions. And you know it takes a long time to develop good questions. … His first use of the question bank was to select questions from it to revise an existing test: T1: We needed a new test because of the changes in the scheme of work … so I looked in the pack and chose the questions that seemed to correspond to our scheme of work, and just pulled a selection out.

Others reported similarly on selecting test items to fit in with their current teaching programmes:

> T4: I used three of them in an end of topic test. … we already have these set up. … But I realised I could stick [some extra questions] in … PT1: I used those as an end of topic test, not just those, but they were included. … as I've looked through, I can see where I can slot some of them in, in that way.

Existing arrangements in some schools meant that it was difficult, or impossible, to use the questions in this way. Because a change in a topic test would have required agreement within her department, teacher T3 considered using the diagnostic questions for short tests to monitor pupil progress during a topic:

> T3: I think at the moment I wouldn't use them as a topic test, mainly because …. We have all departmental data and everyone does the same tests, and we use those to set groups

[2] T indicates a secondary school teacher, and PT a primary school teacher. Numbers are used so that it is clear where different comments cited were made by the same person. I indicates interviewers' comments.

and evaluate pupil progress, and it would go against what's been set up at the moment. …. But I think, maybe, I might use some …. in a short test within the topic.

Another felt constrained by the fact that her school's end of topic tests were deliberately designed to mirror the kinds of questions used in national tests (taken by pupils in England at ages 7, 11 and 14):

I: Do you envisage using any of these questions in the test?
T6: I can't, not now, I don't think, because the tests at the minute are turning into SAT[3] based tests, and it's hard to level[4] these [the diagnostic questions].

This issue of the difference between the diagnostic questions and those used in national tests and public examinations (in particular GCSE, taken at age 16) is one we will return to later. Several teachers, including both those who did and did not use some of the questions in their internal school-based tests, reported using a selection of questions for revision, either prior to an internal test, or in preparation for a national test or public examination.

4.1.2 Pre-topic tests

Several teachers mentioned using the materials to check on pupils' prior ideas before starting on a new topic. Teacher T3 saw her use of one question on particle models of matter:

T3: … as a pre-test, I suppose, in a way. I was using it a bit in that way, to find out what their initial ideas were before we then moved on.

One of the primary school teachers in the group was more explicit about using questions in this way to guide her future teaching:

PT2: I used them near the start of the topic because I wanted to see where the main sort of teaching needed to be at …

Teacher T6 explained how she planned to use some questions on particle models of matter to review pupils' ideas at the end of a lesson. She then went on to discuss how this might be extended to check how well Year 8[5] pupils remembered ideas introduced in the previous school year:

T6: … most of it is going to be a plenary-type thing, checking understanding. … in Year 7 checking understanding after they've done it. And in Year 8 checking understanding before they embark on it.

[3] 'SAT' is the term commonly used by teachers for the national tests taken by pupils in England at ages 11 and 14.
[4] National test results are reported in terms of a pupil's 'level' on a scale from 1 to 8. T6 here means that it is hard to associate performance on a diagnostic question with a particular level of performance on the national test.
[5] Year 8 is the second year of secondary school in England, pupils aged 12-13.

The term 'plenary' here comes from a current national in-service training initiative to improve teaching in the 11-14 age group (the Key Stage 3 National Strategy (DfES, 2002)). One element of this is to encourage science teachers to end every lesson with a 'plenary', or whole-class discussion, drawing together the main points of a lesson and highlighting the key learning points. So the diagnostic materials are here being seen as a means of addressing another current pressure for change in practice.

Checking on pupils' understanding at the beginning or end of a lesson was mentioned by many teachers (see section 4.1.3) as a way in which they had used the questions. None had yet used the materials, however, to construct more formal pre-topic tests, in order to modify their subsequent teaching in the light of this. One reason given was that pupils had not previously met the new ideas they were going to study, and so would not be able to answer questions in a pre-topic test other than by guessing. Where pupils were beginning a new topic that built on ideas they had met previously (sometimes over a year earlier), some teachers (like T6) above mentioned using some diagnostic questions to see how much had been remembered. Another (T10) began a topic on Newtonian mechanics by getting the class to complete a 16-item test that he had been sent for pilot testing during the development of the diagnostic items banks. He was keen, though, that the pupils should not see this as a test:

> T10: As soon as they get it, they look at it and they say, 'oh this is a test'. But I made very clear to them at the start that I wasn't going to mark it, I wasn't interested in … how many they got right or wrong. What I was interested in is what they were thinking, what was going on in their heads. And once they knew that, and they knew they were going to get together and discuss it, then they were quite happy to do it. Not seeing it as a test.

This theme – of using diagnostic questions as a means of leading into a teaching activity – came up in several interviews. It seemed clear that teachers were concerned not only about establishing what their pupils did and did not know, but in linking this to activities designed to take their understanding forward. Thus assessment becomes an integral part of the teaching. This is an issue to which we will return later. It may also be worth noting here that several teachers indicated that pre-topic testing was an way in which they might use the question banks in future, now that they had gained some familiarity with them. And, despite the fact that he had not presented it to his class as a pre-test, teacher T10 was planning, at the time of the interview, to ask his pupils to answer the same questions at the end of the topic, to see how their ideas had changed.

4.1.3 Whole-class discussion

Testing, however, was not the only – or even the main – use of the diagnostic questions. Several teachers commented that, whilst their initial reaction was to see the materials as test items, they quickly recognised that they had other uses, particularly for initiating discussion. As one Head of Department put it:

> T11: I flicked through them and thought, oh yeah, they look like tests, you know, tons of it. And then when you look at it, you think, oh, no, no. And they were so interesting to use.

> The use for me is opening up the discussion, thinking about how they're actually perceiving things, that was the interesting bit.

One use of the diagnostic questions, in every school in the group, was to focus whole class discussion, often at the beginning or end of a lesson:

> T4: For starts and ends, that's been my principal use. … at the start of a lesson, basically testing the ideas taught the previous lesson. You know, testing what I've taught them the lesson before. Often that's how I'll start the next lesson. Have they got it? And sometimes at the end of a lesson.
>
> T6: I imagined them on OHP for open class discussion, that was how I was seeing a lot them when I was looking through. … Come the end, in the last 10 minutes of the lesson, we've done this, that and the other, let's just check and get people to discuss.

In the extract above, teacher T6 is using the term 'discussion' to mean a teacher-led interaction with the whole class – using a diagnostic question to stimulate and focus this. Some others appeared to mean the same when they talked of 'discussion', as in the following extract:

> I: How did you use that in your teaching?
> T4: This was a discussion. After we'd done parallel circuits, after we'd done resistance, there were the two points I wanted to verify that they'd got clear …..
> I: Did the children discuss in small groups or did you do it as a whole class discussion?
> T4: No, I haven't done the small groups, although I had thought when I was thinking about using these that I would do small groups, and I never did. I did it on your own, no discussion … fill it in and then we had open discussion about answers.

This way of using the diagnostic question bank was mentioned by teachers in all the secondary schools in the group. Many clearly found the questions useful as a quick way of focussing the class's attention on a key point that had been taught, or to get them to think about a point that was about to be explored.

4.1.4 Small-group discussion

Whereas teacher T4 above had considered using the materials for small-group discussion, but not actually done so, many in the group found that one of their preferred ways of using the diagnostic materials was to stimulate discussion in pupil groups. The extent to which the diagnostic materials encouraged discussion-based activities was greater than anticipated, and the positive evaluation of this discussion by many of the teachers was one of the most encouraging outcomes of this work. Teachers used different ways of organising such discussions:

> T1: I put them in groups of two … and I asked them … to look at the questions, read them carefully and then try to find the appropriate answers. But I gave them only two questions at a time. … and I gave them maybe initially about five minutes on each and then I asked around the room for their answers …. and when we'd been round the room discussing these, then I clarified what I thought the question was after at that point and then they turned on to the next page.

T3: One thing … that I really like is group discussion. … So I photocopied the questions in little booklets and got them to discuss in pairs and then in a four …..

T10: I asked them to do the questions individually, and they sat down and did them on their own. And then I put them into groups of four and I said 'I want you to compare your answers. If you agree, fine, if you don't, then I want you to sort out between the four of you what you think the right answer is.' And we did that in batches of, I don't know, half a dozen questions. … And then having gone round and seen what they were doing, I picked out the ones that were causing the most discussions in the groups. Once the groups had discussed and agreed within the group, what they thought the correct answer was, I then asked different groups to tell the other groups what they thought the right answer was. And we had a whole class discussion about what the right answer was.

T11: I did it in pairs with that group … and they had to come up with a paired decision as to the answers. …. I'd done the others individually and then gone through them and I thought, oh I'll just see. And I thought it worked quite well in pairs, ... especially when you've got to make a decision and some of them aren't obvious. So I only did that just to vary what I'd done, but I would use it again.

Teachers' views on the quality of these discussions are explored more fully in section 4.2.2 of this report.

4.1.5 New activities

A few examples were given of new lesson activities based on ideas from the diagnostic questions. One example described was of a poster-making activity for pupil groups, on atoms and molecules, using coloured paper circles:

T5: … we were talking about metals and we were going to talk about metal compounds. So I selected questions … about elements, compounds and mixtures. And I did a thing like that [pointing to display board]. We got atoms – lots of different colour circles – and we talked about it just very, very briefly. And then I asked somebody to say, … put an atom down. Make a molecule. And then noticing the difference between some molecules which are made of two different things and elements which are made of only one thing … It's the first time I've ever done it in that way, and that was as a result of this. … even with the bottom group, it worked so much better. I mean, they understood a lot better, this Year 9, so that was brilliant.

Another teacher (T8) had a sixth-form pupil modify a group of paper-based questions on distance, time and speed into a computer-based animation, which was then incorporated into the school's teaching module on forces and motion for Year 9 (age 14) pupils. Several others suggested that computer-based versions of some questions could be useful for individual work, particularly if automatic feedback could be provided. Most, however, felt that interaction in small groups and verbalisation of ideas were important in using the diagnostic questions to promote learning.

4.1.6 Teacher development

In one school, involvement with the project led to an unplanned use of the diagnostic materials – for developing teachers' own understanding of the science ideas they have to teach. This arose from the department's participation in trials of questions during the development of the item banks. The starting point was a concern about the performance of their pupils on some questions on forces and motion – a topic which they felt they had taught well. The teachers in the department, a mixed group of biology, chemistry and physics specialists, decided to use a staff training day to do the same test themselves, individually, and then discuss their answers. This proved very revealing, as the Head of Department reported:

T2: There were certain questions staff were getting wrong. You know, this member of staff got that one wrong there, they got that wrong there, they got that one wrong. And this is somebody who's been teaching for the best part of 30 years. ….One member of staff said, 'this has been really good for me … I got that wrong, I clearly had a complete misconception about what actually was going on here.' And because she'd got a couple of the physicists there, who explained it to her, she said 'that's cleared it up for me now it's just clicked, it clicked it into place for me.'

I: Teachers finding out that they've got things wrong, they could get rather defensive about it ….

T2: No, the comment at the end of the day was 'this is the best piece of INSET I've had in two years.'

In addition to clarifying these teachers' own understanding of some basic points in Newtonian mechanics, the discussion stimulated their thinking about how pupils learn this topic (and science more generally):

T2: … the whole day, it stimulated debate like I've never seen before in terms of the ways in which you understand things, and how you grasp things and how you learn things, really. Which, for my department, was something really good. Because, you know, we've never had time to sit down and actually try and think about how children learn. Some of the younger staff were saying, 'well, you know, how do you learn this? I can't remember when I actually suddenly grasped this.' They felt that some of these questions were actually getting you to the point where it would allow it just to twig[6] for you really.

This school department subsequently rewrote their teaching unit on this topic for Year 9 (age 14), incorporating several of the diagnostic questions from the bank in their revised scheme. Four teachers were interviewed after they had taught this new scheme, and their views on the diagnostic question banks are amongst those reported in the discussion below.

[6] Colloquial. T2 is expressing the view that answering these questions, in sequence, can lead you to an understanding of the ideas, rather like Socratic questioning.

4.2 Teachers' views on the diagnostic questions

The previous section has described the main ways in which teachers chose to use the diagnostic materials provided. In the interviews, we also explored teachers' views of the qualities of the diagnostic questions, and of how well they worked in practice. Comments fell into two main groups: on eliciting learner's ideas and thinking, and on stimulating discussion.

4.2.1 Eliciting learners' ideas

Many teachers mentioned the value of the diagnostic questions in eliciting learners' ideas, and in particular identifying 'misconceptions'. For one teacher, the fact that they were quick to use was important:

> T4: When I [first saw these], I thought, they're absolutely superb for attacking the misconceptions and misunderstandings and quickly finding out what they're really thinking, very quickly finding out what they're really thinking. And I thought 'I can use those for that'.

Others emphasised their role in bringing ideas into the open, so that they could be discussed and misconceptions challenged:

> T3: However we teach a topic, pupils always have misconceptions. But if they're more open about them and we can talk about them, then you're likely to challenge the misconceptions and move them forward. So, I think that's what it did, brought them out into the open a bit more.

Several commented that a particular value of the diagnostic questions was that they showed how prevalent certain ideas (including misconceptions) were in a class – contrasting this with the information that a teacher typically obtains from a question-and- answer dialogue with the whole class, in which only a few pupils participate:

> T9: … without the questions, I might never have been aware of how widespread particular misconceptions were.

> T3: I think I was probably expecting some misconceptions ….. But it brought it out … rather than a few individuals that I knew would have misconceptions, they all shared them and were talking about them.

One teacher was particularly attracted by the precision with which the questions address specific ideas:

> T1: … seeing these, immediately I thought, well those are …. unambiguous, which I think is important. … when you're trying to work out what kids can't do, I think.

He went on to contrast the diagnostic questions with questions in textbooks and national examinations, seeing them as much better for discriminating between pupils who have and have not grasped a key idea:

> T1: … with most textbooks, and I've mentioned SATs and GCSE questions, I don't think that the quality of question is always particularly useful in working out what it is kids can and can't do. I don't think they're very sensitive. I think these are very sensitive, they will discriminate really well between kids who really do grasp and don't grasp these ideas.

This comment, and several others, suggests that precise and focused diagnostic questions were seen by these teachers as filling a gap in the resources currently available. One noted 'the simplicity of testing what is often a misconception' as 'one of the things that attracted me to them immediately when I first saw them' (T4). This quality, of helping to identify specific ideas that pupils may hold, made the materials attractive to teacher T6, despite her usual tendency to devise her own teaching materials rather than draw on her department's stock:

> T6: I don't tend to use very many of the sheets we have made up for us anyway. I tend to be the sort of person who creates my own. But having these here, it did make you think, because it does identify strengths and weaknesses quite well, I think. The actual questions themselves, the choices that people are given, I think, do cover misconceptions and so forth, well.

4.2.2 Stimulating discussion

The other quality of the diagnostic questions which was mentioned in many interviews was perhaps less-expected: that they are good for stimulating discussion, in particular well-focused small-group discussion. Teacher T4 expressed a view that emerged in many interviews:

> T4: … so much of what is generated from this is discussion with the pupils … which is what these have prompted a great deal, which wouldn't have been there without them. ... It's prompted more discussions than I would normally have had … which is good.

In some cases, these discussions were long and involved. One teacher (T10) wrote a detailed account of how a 60-minute lesson, for which he had planned a number of linked activities, became an extended discussion of one diagnostic question about the forces on a football, moving along a level pitch after having been kicked. 'The question as to whether or not there was a forward force provoked a heated debate'. After describing some of the points raised in this debate, and how they were resolved, he concluded:

> T10: What I got, from one EPSE question, was an entire unplanned lesson with pupils fully engaged and making real progress with their thinking. Wow!

Some of the teachers in the group felt initially that the questions, many of which involved reading quite a lot of text, would work better with higher ability classes. But several who used the materials with a range of classes reported surprise at the response of less academic sets:

T10: I think what surprised me was how well Set 5[7] took the questions. Because it looked like a test and because it's fairly abstract. But they had a good go at the questions themselves, on their own. And as soon as they were in groups talking about it, the discussion was actually very good. And it was about the questions; it wasn't about what was on the TV last night, or what their friends did yesterday. And they actually got into the discussion, because they disagreed with one another. And one or two of them had some very good ideas and were able to explain things very, very well. With the upper sets, I expected them to want to talk about these things because they're a good group anyway and they do talk about physics and they do ask questions. But the bottom, that Set 5, were talking about it just as well. And they can be quite a difficult group. If … I'm trying to teach something at the board, then that's hard work with them, they don't want to listen to me. But listening to each other, they actually talked about things, I thought, very well.

Several teachers commented that getting pupils to discuss questions in small groups before any whole-class discussion resulted in more pupils expressing their ideas, because responsibility for the answer was shared:

T1: I think it's this whole thing of getting kids to express what they think. This does that because it's all very specific and careful, the way the questions are written. And when you talk to the kids, you know, if a lot of them aren't sure, they just don't want to say anything. Whereas if you have this kind of material, it doesn't make them feel so vulnerable.

T7: Together, as a group, they often put forward ideas from their group. Because that's not me, that's our group saying this, so there's less fear in it. I mean, some have no problem, they'll just say what they want. But this group thing helps, because it's 'we think', not 'I think'.

Another aspect of the diagnostic questions which several teachers thought promoted discussion was the use of multiple-choice formats, where pupils were given several alternative answers to choose between. This led to much more lively, and better focused, discussion than open-ended questions.

T10: I think that's a very useful feature of them, the fact that they do give alternatives, so the kids aren't thinking in a vacuum. They do actually have a starting point.

T13: I do like to discuss but it made it easier. It made a lot more openings for discussion, whereas sometimes with other questions and other ideas, you tend to get limited with what kind of alleys you can go down. … Whereas with those, because … the children had … lots of ideas in front of them. I thought that it was better. … because … they've got ideas already and then they can bring their own in with it.

As a result, the level of pupils' engagement with the questions was high:

[7] Of six sets, ranked in order of previous science attainment.

T13: They enjoyed the lesson immensely to be honest. They really enjoyed it. … there was lot more discussion, you know, when they were working within the small groups and within the class.

Teacher T11 commented similarly that 'the kids did enjoy it … even the ones that they found hard'. Perhaps one of the less expected outcomes of this work was that many pupils, even those in less academic sets, can be interested in discussing quite abstract ideas, and arguing about quite fine distinctions, if these are presented in a way that facilitates participation, by providing a range of ideas to consider and debate, and encourages peer interaction and collaboration in reaching an answer.

4.3 Impact on practice

Having looked at how teachers used the diagnostic materials provided, and their views on them, we now turn to the central question of this research project: in what ways, and to what extent, does the provision of diagnostic assessment resources lead to changes in teachers' practices? At the surface level, providing these diagnostic questions clearly did lead to changes in teachers' actions. Every teacher interviewed had used some of the diagnostic questions in their teaching – thus adding new activities that they had not previously included in their teaching of this topic. Of more interest, and long-term importance, is evidence of more fundamental change, in teaching sequences, or learning objectives, or teaching methods and approaches. In one school, data on their pupils' understanding of basic ideas led directly to a review of how one science topic was taught. The Head of Department's comments make clear that that the fact that these data came *from their own pupils* was important in stimulating change:

T2: … we'd got some feedback that really, the methods that we were using were not promoting good understanding of the topic. … You know, we'd got [data on] schools P,Q,R and S, but [my colleagues] were interested in school P, which was us. … They weren't so bothered about what other schools got. …. You know, it was, hang on a minute, why is our school not doing better?

This was the only example in the group of a complete revision of a teaching sequence – though this is not surprising as the teachers only obtained the materials at the beginning of the school year, when plans for teaching would already be in place. Several, however, planned to incorporate activities based on the diagnostic questions into their department's scheme of work for the next school year. For teacher T4, it was almost self-evident that this would happen: 'Of course, these will now go into the Scheme of Work, without a doubt. Yes.' (T4) Another discussed his plans in more detail:

T10: We're actually now starting to use the Key Stage 3 electricity ones and forces and motion ones, because we're rewriting the Scheme of Work for next year. … [Named colleague], who's the other physicist in the department, she's writing the electricity topics, so she's got the diagnostic questions for the electricity topic. And we're debating at the moment how to use them. One thing we're thinking of doing with them is using an electronic version through the data projector and using those as lesson starters, for example. So we're hoping to fit those in now, into the Schemes of Work.

As Schemes of Work at Key Stage 3 level (pupil age 11-14) are usually developed collaboratively within school science departments and then used by all staff, they are an important vehicle for dissemination of ideas, materials and approaches within a school.

Several of the group gave examples of unexpected findings from using the diagnostic questions with their classes. For example, one primary teacher was surprised that many of her pupils gave wrong answers to a question about opening a switch placed between two batteries in a series circuit with a lamp:

> PT2: It's easy to take for granted that by that age the children have got a fairly good idea of the complete circuit. But I found there was one question in particular, that quite a significant number of children got wrong … which surprised me. … I think what they thought was that, well it's connected together, therefore it must work, because it, you know, it goes right across. … What it showed me was … that I still needed to consolidate the work on the complete circuit, which surprised me slightly.

A secondary teacher talked of her surprise at her pupils' answers to a question about whether some given events are examples of physical or chemical change, and whether they involve 'new substances' being formed:

> T11: I did one [question], to start me off, thinking, 'oh we'll do that, that looks easy' … I used that one, and I couldn't believe them, they were rubbish. I thought, 'well, they'll all know that'. And they wrote absolute nonsense.

More positive unexpected outcomes were, however, also reported. Teacher T9 concluded that his class 'understand speed/time graphs better than distance/time graphs.' For him, using the diagnostic questions:

> T9: … did make me think, 'Oh actually, they do understand it'. So they gave me a bit of an inner glow. Perhaps I gleaned from that the fact that, perhaps with some of these groups, I should push it along a bit quicker.

Overall, however, there was a stronger sense that the pupils' answers confirmed expectations than provided surprises. Several teachers commented that the diagnostic questions highlighted pupil difficulties that they already knew of. For teacher T16, the effect of this was to make her question the pace of instruction:

> I: Did you feel that you learned anything from them that you wouldn't have expected to pick up anyway, about pupils' understanding?
> T16: I think it confirmed, you know, the areas that I knew they struggled with. But I think it sort of made me more aware that, you know, perhaps I thought they had understood it through my teaching. And it made me actually more aware that, no, I had to stop and revisit certain areas. Even though I knew those were going to be the harder bits, you know, some of the things that it did pull up, it made it very clear that the students hadn't understood where perhaps I thought they had, and I was moving on. So it was very good in that respect.

Teacher T4 also felt that the diagnostic questions elicited misconceptions he already knew about, from past experience. The impact on his practice was more in terms of *how* he addressed these:

I: So the feedback that you got from the pupils, did that have any influence on how you taught the topic as compared to how you would normally teach it

T4: (pause) Not really because I know that misconception's there. I've known for years that misconception's there so I always address it, I always attack that misconception in some form or another. This was just a way of doing it. … I've been teaching the topic for so long now, you tend to register the things that they commonly misunderstand or get wrong. So I've had some means of attacking those things. These are the quickest and best way I've ever seen of doing it.

He did not envisage any major changes in the way he approached this topic, electric circuits, but felt that the materials had enhanced his teaching of this topic, and would use them again:

T4: I was very, very happy with the way it went this time. [Next time] I might do more overheads. I only did two, but then only had time to actually use one of them. I'll probably use the two, actually, with Year 10.

When asked directly, however, he acknowledged that using the questions had altered the way he taught:

I: So would you say that using these questions is influencing your own teaching style?

T4: Oh it has, without question. In a beneficial way. … I mean if I was the sort of teacher that was always prompting discussion then it probably wouldn't have been a necessity, I wouldn't have needed that. But I did need that and it's helped, without question, it's helped. I'm having more discussions in class than previously, which is a good thing.

Several in the group made similar comments, about how the diagnostic questions had led them to make greater use of discussion, including pupil-pupil discussion in small groups, in their classes. One teacher, a chemistry graduate in her first year of teaching, discussed how the materials helped her to develop ideas in discussion with the class, rather than having to tell the pupils the key ideas:

T13: I'm not very experienced and, especially with forces and things, when I bring it together at the end, or when I'm trying to bring it in, the fact when you have the force opposing motion, what would happen to the speed and so on – I tend to find that I more or less give it to them in a way, because trying to get them to discuss it, to bring it in, I find it quite hard. Whereas it was done for us, for them to discuss. And I was able to do the diagrams on the board and they came up and did the forces and arrows and things.

I: Right. So rather than you having to give them the answers, or tell them things ....

T13: We came to it together, which is a lot better. I thought like I'd actually done proper teaching then.

A Head of Chemistry with several years' teaching experience, talked similarly about how the diagnostic questions on forces helped her teach this topic differently, and in a manner she found more satisfying:

T15: I'm not a traditional teacher in my approach, it's not my style of teaching. But I've probably stuck a bit more to that in physics because of not being as confident in the subject. Whereas this has forced me to take a different approach. … I've taken an approach with this that has been much more the approach that I would take with chemistry … much more open, you know, rather than me just giving information and working through things, a much more, sort of, interactive, discursive approach which is a style of teaching I prefer. I think it's a better way of going about things, but perhaps I haven't been as confident in physics before to risk it. So this has given me a bit more impetus to do that, which is good.

I: Is it a matter of confidence or of giving you materials that suggest a way you could do it?

T15: Yes, I think that's right, because I come up with ideas myself for biology and chemistry, but I find it much more difficult to come up with them for the physics area. So yes, for me, it's given me material that I can build things around and have confidence in them working. Which has been nice, because, like I say, it is an area that I find it quite difficult to be imaginative about.

For the two teachers cited above – and several others in the group – providing the diagnostic questions facilitated a change that they were well-disposed to make, but had not previously seen how to. The question packs did not alter their views on how to teach science, but enabled them to implement the form of teaching they favoured in topic areas for which they had not previously been able to do this. Several were quite explicit about the match between new materials and their own preferred 'style' being a significant factor in deciding whether, and how, to make use of them:

T6: … you always have your own, you have things that fit with you, don't you, as a person. There were some things I looked at and thought, yes, I do that already. And some things I looked at and I thought, oh yes, I would like to do that now. It's still me, but it's something I haven't done before. So, yes, I like that sheet. I like that approach. I'll have a go at that.

Similarly, for teacher T3, an attraction of having a large bank of diagnostic materials to choose from was its 'flexibility' of use:

T3: Why were these useful to me? I think they enabled me to use the questions in a flexible way …It fitted in with my own teaching. That's why I chose those questions, because I thought, I can use those in a similar way to that I would teach anyway. It was just like adding to it rather than altering what I was going to do completely.

These perceptions of flexibility of use, and match to preferred teaching style, are clearly significant for uptake of the materials. New materials – including those based on research – are only likely to be used if they 'fit' with current practices and priorities. However, several of the comments cited also suggest that providing materials can stimulate changes that go below the surface level, and are rather more than 'quicker' or 'better' ways of doing the same thing – but involve shifts in emphasis, notably here from more transmission-oriented towards more discursive forms of teaching.

4.4 Teachers' understanding of science

Several of the teachers involved in the project discussed how the diagnostic question banks had helped them, and their school colleagues, to improve their own understanding of the science topics covered. In one school, as discussed earlier, all the science staff tried to answer individually a set of diagnostic questions they had previously given to their pupils, during the trials stage of this project. One teacher at this school discussed this experience, and its effects, as follows:

> T15: I think my understanding of the topic has improved as a result of going through the questions, personally, on a personal level. And I think that, consequently, I will become better at dealing with it in the future, at whatever level it happens to be. Because I think, you know, even science teachers have the same misconceptions. … I had a misconception with forces, that I never knew I had. … we actually did the test thing, as teachers, just to see what we got. And a lot of us had at least one misconception. And when you start addressing that yourself, you start getting more confident. And thinking 'well I'm not [the only one]' I always worry about physics, because I'm not a specialist in it, thinking 'well what if I get it wrong? And I tell the kids something wrong?' Whereas it kind of boosts your own confidence, with the materials, and it also helps you understand it more.

Several others also commented that becoming aware of your own misconceptions in areas you knew to be difficult – and then dealing with them – made you more confident in teaching the topic subsequently. Teacher T13 attributed this to being more able to identify the key learning outcomes, and to anticipate points that would arise later:

> T13: It gave me confidence in delivering it, being able to have looked through it ... Because, I mean, I had a misconception, I found out, when I first looked through them. It was the football one, what forces were acting on the football? I got that wrong. So in a way it was quite good to have all this bank because I felt a lot more confident. It probably showed through, when I was delivering the lesson and discussing it, that I'd seen what was coming up. I knew what responses I was wanting to get …. So I probably delivered it better, I'd seen it all and knew what was there, and what was expected, and what I wanted to get from them, and things like that.

For teacher T11, an experienced Head of Science, answering the questions you were going to ask your pupils to answer was an ideal form of preparation:

> T11: But it's really stimulating to learn yourself, and to know a bit more. And there's nothing better to helping you teach physics than doing one of those little booklets or something similar and getting it all right. Or being taught anything…. It's like normal, you know, the best lesson prep is to answer the questions you're going to give your kids to do.

All the comments above are about teaching physics topics, by teachers whose specialist knowledge is in chemistry or biology. But several also commented on aspects of teaching their specialism. Teacher T1 felt that some questions on identifying forces had required him, and his pupils, to analyse situations more precisely and carefully than before:

T1: When I was working with the Year 9 groups, with the forces questions, that encouraged me to perhaps present certain ideas and concepts in a different way, I think. … What we tended to gloss over in the past is the fact that at any point, well forces are active, yeah? At any point there's going to be more than one. … with these questions, you've got to actually understand more of what the mechanism is, which leads to the forces arising in physical situations. And so there was the whole way the questions developed on those, you know, encouraged them to say, well, there's something happening at this point which means there must be a force acting at that point in that direction and also a force acting at that point in that direction. They're acting at the same point in different directions and one's bigger than the other or whatever, you know. And I don't think that had ever been examined in detail in any exams that I've seen. And so we needed to discuss that.

I: So you think that has changed how you teach the topic?

T1: Yes, definitely. I think that's quite a significant difference, as well.

Whilst not changing teacher T1's physics knowledge, the questions were stimulating a rethinking – which he regarded as 'significant' – in how he conceptualised the topic from the perspective of communicating it to learners – his pedagogic content knowledge (PCK) (Shulman, 1987). In the following extract, a very experienced physics teacher reflects on how a discussion, in a class of 15 year olds, of some questions about the motion of objects that have been set in motion by a kick or throw, made him aware of the value of the concept of 'momentum' for clarifying pupils' ideas and helping them towards an understanding of motion. Again this can be seen as a development of his PCK, as a result of working with the diagnostic questions:

T10: Although momentum isn't on the syllabus, during the discussion it was perfectly obvious that we needed some sort of word to describe the property of a moving body, independent of the forces on it. So we just invented the word 'momentum' as a way of talking about things. And the fact that a force always involved two objects – it's always a force of one object on another – was very, I mean, as soon as we got that, and as I say it came out of the discussion, it became necessary in the discussion to use momentum. I had no intention of using momentum. But once we got that idea introduced, as soon as we looked at this one, I mean, kids who would have normally have said, 'well it's moving up, so there must be a force upwards', immediately said 'no, no, no, there's only one force and that's a downward force and that's why it slows down.' So I felt that they had managed to [grasp something important]. … Momentum used to be on the GCSE syllabuses, and I was never happy with it being there really. I think it's because all the questions were about conservation of momentum, you know, like explosions or collisions, and it was all quantitative and there was no notion that this was actually a useful concept to help you understand the difference between forces and motion. I think that qualitative notion of what a force is, and what momentum is, ought to figure more strongly in the syllabus than it does.

In addition to these very specific and detailed examples, several other teachers in the group talked in more general terms about the usefulness of the diagnostic questions in clarifying learning outcomes in the topic areas covered, highlighting the most important outcomes, and providing a clearer indication than syllabus statements of what pupils might be expected to be able to do, if they have successfully grasped these ideas.

4.5 Impact on pupil learning

Changing teachers' practices is not an end in itself. The end goal is to improve learning outcomes, either in terms of pupils' understanding or enjoyment of the subject. It was, however, difficult to obtain a 'hard' measure of the impact on learners of the diagnostic materials. The teachers involved in the study made changes – large or small – in the way they taught one topic on the syllabus. The data available on learning outcomes from groups taught in previous years, which might be seen as a baseline, varied in quantity and nature from school to school. Also, the diagnostic questions encouraged a focus on understanding of basic key ideas in the topics, which many teachers in the group recognised as different from that of national tests (which many of their internal school tests mimicked, in order to give pupils practice in tackling such questions).

When asked directly about pupil learning of the key points probed by the question banks, teachers invariably said that they felt these were better. The following two comments are illustrative of the kinds of responses given:

I: Do you think their answers were better than you would have expected in the past?
T5: Much better. They understood a lot better, this Year 9, so that was brilliant. … the marks were very good. … Apart from an odd one or two, … everybody was able to understand much, much easier, and was able to answer those questions.

I: Did you feel they learnt the ideas you wanted them to learn?
T13: I think they did, from my point of view. I think they took it more on board, because they had to discuss it, and they had to back up their answers, and they had to say why they thought it. And when they were saying all of it, I felt quite confident they'd got it and they knew why. It wasn't just a case of just reeling it off and writing the right answer. You know, because they've actually thought about it.

Teacher T15 felt that many pupils in her group would have grasped the key ideas anyhow, but that the teaching approach which the diagnostic questions encouraged had led to better learning by the less able pupils in the group. As the new approach had been more enjoyable for the able pupils, they had also gained from it:

T15: I think the majority of them would have got the ideas anyway. A lot of them would have. The less able students in the group were forced to think about it when perhaps they would otherwise have just written, because they're the sort of group that will write things down and take notes and keep themselves busy, and not necessarily be understanding all the time. And this sort of forced them to think, some of the less able ones, who would normally perhaps rely on someone else, to think for themselves, which helps. The fact that we had discussions going meant that that happened. My feelings are that certainly the less able in the group were much better informed about the topic at the end of the process than they were at the beginning and that it was helpful in that respect. … based on the evidence of the questions that they've answered, and now looking at the past SATs questions that we're looking at, topics that are related, how they're answering those now, and they're doing a better job of them. I think the more able students in the group enjoyed it, probably would have got it anyway, but enjoyed it. And therefore, that was as useful as anything, you know, in a way.

Some teachers in the group saw similarities between the style of the diagnostic questions and those used in national tests (SATs). One primary school teacher, involved in marking national tests, saw both as probing understanding rather than recall:

> PT1: I thought these related more to the way the SATs are going. The other tests I've used have been like a regurgitation of what they've done in class. Now this, … wasn't exactly the same. Understanding, rather than regurgitation.

Rather more in the group, however, tended to note differences between the two types of question:

> I: Do you think those questions are like SATs questions?
> T12: Not really, no. One or two of them have bits in them that you might imagine seeing on some of the SATs questions. But it was because they were quite different from SATs questions that I think the kids got more enjoyment out of doing those than they would a normal, sort of, SATs-derived test.
> I: So what did you think the difference was?
> T12: … sort of, challenging thinking, whereas the multiple-choice questions in the SATs tend to be a bit more straightforward.

Where teachers saw differences between the two types of question, they tended to prefer (and often to feel their pupils preferred) the diagnostic ones:

> T9: I'm waiting for the day that we actually have some exam questions like these. They would be good for exam questions, but we very rarely see them in Key Stage 3 or Key Stage 4 in fact. But I still insist on teaching them.
> I: Why do you think they'd be good as exam questions?
> T9: Well, for a start there's only one right answer. Secondly, they're probing into the understanding, of current flow in series, that a lot of kids [find difficult]. … It's going through a bulb, so some of the electricity must be being used up … I like questions like this because it probes into their understanding of current flow. I'm just surprised that we don't get more of them, to be honest.

However, a perception that diagnostic questions probe different kinds of understandings from national test questions might restrict their uptake and use. This issue arose in several teacher interviews. Teacher T11 argued that a good understanding of basic ideas – and practice in selecting and discussing explanations for phenomena – would lead to better national test performance:

> T11: I can remember thinking, although they're not the same as SATs questions, if they can do this and get to grips with this, this'll be really useful for being able to give your explanations and things in national tests. I can remember thinking that, even though they're not the same. They're not the same sort of questions.

Teacher T10 took a similar view, also arguing that teaching for understanding was more useful to pupils in the longer run:

T10: … it's the old chestnut, you either teach kids to the test, in which case they tend to do fairly well in the test, or you actually try and get them to understand what's going on, in which case they do reasonably well in the test, one would hope. And the second alternative strikes me as a better alternative in terms of teaching them some science, … because I think you're more likely to help them understand the world around them. … If they actually understand, it's going to be more use to them than simply being able to answer a SATs question. Who wants to answer a SATs question in real life?

These comments do, however, highlight a fundamental issue regarding evidence-based practice in education. This is grounded in the view that different teaching approaches can be compared experimentally to see which 'works' best (Davies, 1999). Two or more teaching approaches to the same topic can, however, only be compared if there is agreement about the learning outcomes, not simply at a general level but at the detailed level needed to develop and agree a common outcome measure. This is rarely the case in practice.

This perceived difference in style and purpose between diagnostic questions and those used in 'high stakes' external tests may have been one reason why many teachers in the group chose to use the questions *during* the teaching of a topic, rather than for summative assessment at the end. Several commented that they found out more about their pupils' ideas by observing their response to activities based on the diagnostic questions – and in particular listening to how they expressed ideas in their own words:

T3: I actually felt I'd got more of an idea going round and eavesdropping on their conversations, than I did with their written comments. The written comments were of some use, but I think it was the discussion and listening to that that I found more useful. I think they were only writing down what they definitely, definitely thought. So some of these ideas that were greyer, I don't think they wanted to write those down.

T10: … the traditional assessment method, I don't see that as the main use. Because all the things that count in terms of assessment, nationally, are SAT-style questions. And those aren't SAT-style and GCSE-style questions. So in terms of formal assessment, if we want any kind of predictive power, in terms of what kids are going to do in the SATs or GCSE, then we just use SATs questions. Those [the diagnostic questions], to me, they're much more useful for getting them thinking. The electricity ones, for example, you know, the two-part questions. And the way you get at kids' thinking is to talk to them about it, I think. And get them to talk about it.

If lesson activities provide good opportunities to assess pupils' understanding and to monitor their progress, then further information from an end-of-topic test is unnecessary, as teacher PT2 explained:

I: So you didn't use any of the questions in an end of topic test?
PT2: No. I didn't do that because, in a sense, by using them the way I did, I'd found out what I needed to know.

5 Outcomes and implications

The aim of this project was to explore the extent to which science teachers' practice might be changed – and become more research evidence-informed – by providing them with a bank of diagnostic questions and encouraging them to use these in the course of their own teaching. So does this work? Is it a viable means of increasing the impact of research findings on practice? Any answers offered to these questions must be tentative, as the study was small in scale, involving just 10 schools and around 20 teachers. Several of the teachers involved in the project had previously helped in the development of the question banks. The others had all volunteered to take part, some after seeing or hearing a presentation about the project at a conference or teaching workshop. So they were disposed from the outset to use them – and the fact that we indicated we would keep in regular contact to hear how they were using them must also have acted as an incentive to 'do something' with them.

Nonetheless, these case studies strongly suggest that teachers' practice can be significantly influenced by making available teaching materials based on research findings and insights. The practices of this group of teachers were significantly modified (and, in their own view and that of the researchers, enhanced) by access to banks of diagnostic questions informed by researchers' analyses of content and experience in probing understanding, and by the findings of research on science concept learning. These helped teachers to identify more precisely, and to focus teaching more strongly on, the key ideas that are at the heart of an understanding of these science topics, and which provide a basis for further learning. In the words of one teacher, 'they remind you of what you should be doing, when there are so many other things that can distract you' (T12). Teachers valued the way in which the questions enabled them quickly to assess the understanding of all the students in a class, rather than sampling a few individuals. The view that structured diagnostic questions are particularly useful for stimulating small-group and whole-class discussion was expressed very widely within the group. These discussions were often characterised by high levels of student engagement, and lively debate about ideas and explanations, which provided clear evidence of student learning. Several teachers also indicated that the diagnostic materials helped them to teach science topics outside their specialist area in more interactive ways, and with a clearer understanding of which ideas to emphasise.

This is not, however, to claim that *any* teaching materials claiming to be based on research would have influenced teachers' practices. It was clear that these teachers were making judgments about the qualities of the materials, both in terms of presentation and (more importantly) of match to their teaching situation and the challenges they are facing. Indeed teachers' initial reactions to the diagnostic question banks were invariably based on their judgement of its usefulness as teaching material, rather than on the fact that it is (or claims to be) research evidence-informed. Fullan (1991) identifies 'quality and practicality' of a new approach as one of the main factors that influences its uptake, and hence the extent to which educational change ensues. Similarly, Doyle and Ponder (1977-8) suggest that new materials and approaches must pass the test of teachers' 'practicality ethic' to be taken up, that is, they must be seen to be salient to their needs, fit their situation and context, and it must be clear how they can or should be used. These diagnostic materials seemed to be seen by teachers as 'practical' in this sense. First, they fitted well with other current initiatives. We have already discussed how several teachers used the questions to address other current demands, such as the emphasis on whole-class teaching at

the beginning and end of lessons ('plenaries') in the Key Stage 3 National Strategy for Science (DfES, 2002). Second, they had a low 'entry threshold': they could be introduced initially in relatively small ways, to 'test the water', without great risk or time commitment. Third, they were seen as a 'flexible' resource, that could be used in ways that fit with the teacher's own preferred methods of working. Specific questions could be chosen to fit the current needs of pupils, schemes of work and syllabuses. Compared to a fully worked-out teaching sequence, these diagnostic question banks left many choices and decisions to the teacher. Indeed the banks deliberately included more questions on many topics than teachers could possibly use, so that some selection and choice was essential. So, in using them, teachers have to exercise considerable professional judgment, giving them a stronger sense of 'ownership' of the teaching programme that emerges. Ogborn (2002) notes that there is inevitably an element of 'transformation' when teachers implement fully worked out teaching schemes and programmes; here it was apparent that this was intended. Flexibility of use, however, carries with it a risk – of relatively minor and superficial change, involving improvement or enhancement of lessons already planned, rather than a more thorough review of teaching schemes and methods. In fact, given that we were monitoring these teachers' use of these materials 'first time around', the extent to which they were used, and the variety of ways in which they were used, were quite striking. Several, as noted earlier, planned to incorporate a selection of the materials when next revising the departmental scheme of work for these topics.

In part teachers' generally positive response to the materials may be because they offer a tangible resource that can be used directly. Harland and Kinder (1997) suggest that, of the range of possible intended outcomes of teachers' continuing professional development (CPD), the easiest to achieve are 'material and provisionary' outcomes – where the aim is to make new materials available to the teacher. At the other end of the spectrum, the hardest to achieve are 'value congruence' outcomes – where the aim is a change in the teacher's values towards those implicit in the new approach we want to encourage. At first sight, the adoption of a more interactive and discursive teaching approach, in which assessment is an integral part of teaching, might be thought difficult to achieve, and to require a considerable value shift for many teachers. This study suggests, however, that the barrier is at the practical level of seeing *how* to make the change, rather than the level of values. The diagnostic questions enabled teachers to teach these topics in ways they favoured, but had not previously seen how to put into practice.

We have already commented that most teachers' initial response to the materials was based on a judgment of their usefulness as teaching resources – that they were materials they could imagine working well with their classes. In fact this perception of the quality of the questions as a basis for teaching activities was a stronger influence on teachers' actions than the evidence of pupil learning they obtained from using them. This contrasts somewhat with the effects of the Force Concept Inventory (Hestenes et al., 1990) discussed earlier, where university teachers' perceptions of their students' misconceptions were a major trigger to change their teaching approach. The reasons for this apparent difference may be worth exploring briefly. The teachers involved in this project did not appear to find their pupils' misconceptions – or the fact that some were very prevalent and hard to change – surprising. With a few exceptions (discussed earlier), most of the misconceptions and learning difficulties revealed were ones the teachers already knew about and anticipated. Teaching science to the whole population, rather than to a self-selected group studying tertiary level physics, school teachers seemed more willing to accept that

levels of understanding of these ideas were likely to be limited. Whilst interested in 'better' or 'quicker' ways of checking understanding, a majority saw the diagnostic materials primarily in terms of lesson activities which focused pupils' attention on important ideas, in ways which promoted lively on-task discussion, which pupils seemed to enjoy.

Teachers' understandable concern with the inter-personal dynamics of lessons may result in a focus on *activities* that 'feel' worthwhile, rather than on *learning outcomes*. In these case studies, it was rare for a teacher to state clearly the main learning outcomes of a sequence of lessons, and much more usual for a teaching unit to be characterised in terms of the topics covered or the activities involved. One teacher, for example, when asked about the objectives of a teaching unit on electricity, said it was 'about energy transfers using electricity, and introduces the ideas of resistance and voltage'. This level of detail was typical. A pre-requisite for diagnostic assessment, however, is clarity about learning objectives. And unless the proportion of pupils attaining certain key objectives is seen by a teacher as an important indicator of 'success', collecting diagnostic data is unlikely to lead to significant change in practice. Science education researchers may attach higher priority to understanding of key ideas that emerge from a careful analysis of the knowledge structure of a domain, than do teachers whose more pressing daily challenge is finding ways to keep young people interested in, and productively engaged with, science. It is not surprising, therefore, that teachers' responses to these diagnostic materials reflected their conceptions of teaching as well as their conceptions of subject matter.

As explained at the beginning of this report, one reason for the approach adopted in this project was the evidence, as reviewed by Black and Wiliam (1998a, b) that the use of formative assessment can lead to significant learning gains. Their subsequent work to help teachers put this into practice suggests that this requires significant changes in pedagogy (Black et al., 2002). This project suggests that provision of diagnostic materials leads to changes in lessons which teachers see as improvements, but that additional training and support – and perhaps, further development of the materials[8] – may be needed if teachers are to use the materials for more systematic formative assessment. These diagnostic materials seem to be quite readily *assimilated* into current practice, but often in ways that involve little *accommodation* of that practice to incorporate different, or additional, methods and approaches.

In one school, where involvement with the project stimulated a major revision of the department's teaching approach to one topic, the stimulus was data on their pupils' performance on a given set of questions. This tends to corroborate a hypothesis which is implicit in the design of the project, that data collected *in your own context* is a more powerful stimulus to change than reports of data collected in other settings – even if these are from much larger samples. What seems to matter is 'relatability' of research findings to your own context (Bassey, 1981). This might seem to argue for teacher action research as the engine of improvement. In fact, we would suggest that what is needed is a collaboration between researcher and teacher. The choice of questions used to evaluate current practice is crucial, and determines what happens thereafter. Developing well-focused diagnostic questions is far from easy. It requires a good knowledge of the research literature, and a well-developed 'map' of the structure of knowledge in the domain

[8] We are continuing to work with teachers to develop short, focused 'tests' on key ideas, which can be used to give a quick measure of pupils' understanding, and exploring the impact of these on practice.

in question, which many teachers – even quite experienced ones – do not possess. A productive role for the researcher, then, may be to 'translate' research findings into materials that steer the teacher's data collection efforts, and hence their teaching, in productive directions.

The teachers involved in this project seemed to accept that some ideas in science are difficult for learners, and did not expect research to provide a solution that would transform this situation dramatically. On the other hand, they did not therefore seem to conclude that improvement of practice (and learning outcome) is impossible. Changes could lead to more modest gains in both understanding and interest. By 'translating' the findings and insights of research into teaching materials and guidance that teachers can immediately use, and accepting that it is more productive to aim to 'shape' rather than to 'control' teachers' actions, we may be able to enhance significantly the interface between research and practice.

Bibliography

APU (Assessment of Performance Unit) (1988-9). *Science at age 11. A review of APU survey findings 1980-84* (and similar volumes on *Science at age 13,* and *Science at age 15*). London: HMSO.

Bassey, M. (1981). Pedagogic research: On the relative merits of search for generalisation and study of single events. *Oxford Review of Education,* 7(1), 73- 94.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Black, P., & Wiliam, D. (1998b). *Inside the Black Box: Raising standards through classroom assessment.* London: School of Education, King's College.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the Black Box: Assessment for learning in the classroom.* London: School of Education, King's College.

CLISP (Children's Learning in Science Project) (1987). *Approaches to teaching the particulate nature of matter*. Leeds: Centre for Studies in Science and Mathematics Education.

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47 (2), 108-121.

DfEE/QCA (Department for Education and Employment/Qualifications and Curriculum Authority) (1999). *Science: The National Curriculum for England*. London: DfEE

DfES (Department for Education and Skills) (2002). *Key Stage 3 National Strategy. Framework for teaching science: Years 7, 8 and 9.* London: DfES.

Doyle, W., & Ponder, G. (1977-8). The practicality ethic in teacher decision making. *Interchange,* 8(3), 1-12.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. London: Routledge.

Fullan, M. (1991). *The new meaning of educational change.* London: Cassell.

Harland, J., & Kinder, K. (1997). Teachers' continuing professional development: Framing a model of outcomes. *British Journal of In-service Education*, 23(1), 71- 84.

Haslam, F., & Treagust, D.F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration using a two-tier multiple choice instrument. *Journal of Biological Education,* 21(3), 203-11.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30, 141-157.

Hillage, J., Pearson, R., Anderson, A., & Tamkin, P. (1998). *Excellence in research on schools.* Research Report RR74. London: Department for Education and Employment.

Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity.* Utrecht: CD-ß Press.

Lijnse, P. (2000). Didactics of science: The forgotten dimension in science education research? In R. Millar, J. Leach & J. Osborne (Eds.), *Improving science education: The contribution of research* (pp. 308-326). Buckingham: Open University Press.

Mazur, E. (1997). *Peer instruction.* New York: Prentice Hall.

Ogborn, J. (2002). Ownership and transformation: Teachers using curriculum innovation. *Physics Education*, 37 (2), 142-6.

Pfundt, H., & Duit, R. (1994). *Bibliography: Students' alternative frameworks and science education (4th edn.).* Kiel: Institüt für die Pädagogik der Naturwissenschaften (IPN).

Redish, E. (2003). *Physics education.* New York: Wiley.

Shipstone, D.M., von Rhöneck, C., Jung, W., Kärrqvist, C., Dupin, J.-J., Joshua, S., & Licht, P. (1988). A study of students' understanding of electricity in five European countries. *International Journal of Science Education*, 10(3), 303-16.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57 (1), 1-22.

Tiberghien, A. (2000). *Designing teaching situations in the secondary school.* In R. Millar, J. Leach & J. Osborne (Editors): *Improving science education: The contribution of research* (pp. 27-47). Buckingham, UK: Open University Press.

Treagust, D.F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-69.

Viennot, L. (2001). *Reasoning in physics.* Dordrecht: Kluwer.

Viennot, L., & Rainson, S. (1999). Design and evaluation of a research based teaching sequence: The superposition of electric fields. *International Journal of Science Education*, 21(1), 1-16.

White, R., & Gunstone, R. (1992). *Probing understanding.* London: Falmer.