

# **Evaluating Inquiry-Based Science Developments**

**A PAPER COMMISSIONED BY THE NATIONAL RESEARCH COUNCIL IN  
PREPARATION FOR A MEETING ON THE STATUS OF EVALUATION OF  
INQUIRY-BASED SCIENCE EDUCATION**

**By Dr. Wynne Harlen  
University of Cambridge and  
The University of Bristol**

**11 May 2004**

NOTE: This paper was commissioned as background for a meeting on May 11, 2004 titled, "Evaluation of Inquiry-based Science". The meeting was supported with funding from the Bristol-Myers Squibb Foundation. Opinions and statements included in the draft papers are solely those of the individual author, and are not necessarily adopted or endorsed or verified as accurate by the Board on Science Education or the National Academy of Science, including the National Academy of Engineering, Institute of Medicine, or National Research Council. The National Academies own copyright of the paper until August 31, 2004 and will release copyright assignment back to the author on or about August 31, 2004. Permission to use would then need to be sought from the author.

Copyright © 2004 National Academy of Sciences. All rights reserved. No part of these pages, either text or image may be used for any purpose other than personal use. Reproduction, modification, storage in a retrieval system or retransmission, in any form or by any means, electronic, mechanical or otherwise, for reasons other than personal use, is strictly prohibited without prior written permission.

# Evaluating Inquiry-Based Science Developments

## Executive Summary

### *Introduction*

This paper is written to provide an introduction to discussions of what can be learned from existing research and experience relating to the effects of inquiry-based teaching and learning. It considers the meaning of inquiry and of inquiry-based instruction and the obstacles to implementation and evaluation, some of which follow from the nature of the changes in teaching that are required. Some case studies of evaluations are used to illustrate factors that need to be taken into account, and from these are drawn implications for future evaluations.

### *The nature of inquiry-based teaching and learning*

Inquiry-based learning in science means that students develop understanding through using mental and physical skills to gather evidence about the natural and made world. This way of learning is consistent with the modern view of the nature of scientific activity and of how learning takes place. Learning through inquiry not only means that students learn with understanding, so that their knowledge is applicable, but also that they learn about learning. Both of these are important outcomes of education if future citizens are to be scientifically literate.

When inquiry-based teaching is practiced, teachers and students are involved in well-defined actions, which differ in several respects from current classroom practice. This paper argues that interactions among students and between students and teachers are needed for inquiry-based learning, with the teacher having a key role. Thus bringing about the required change in students' experiences is a two-step process, in which teachers' understanding of the changes needed is the first step and the provision of opportunities for students' learning is the second. The main avenues for bringing about change are through the professional development (PD) of teachers, or the provision of classroom materials, or a combination of these. When change is attempted through these inputs, there are numerous other factors that impact on teachers, on teaching and on students, that act to dilute – and in some cases cancel – their effect. Thus there is a danger that when inquiry-based inputs are evaluated, the students' experiences may not be as intended.

### ***Issues of evaluation design***

Unless the intended learning is being experienced by students there seems little point in assessing learning outcomes. However, even when this is the case there remain a number of issues concerning the evaluation process. The chief ones concern: the length of time during which the inquiry-based learning has been in operation before it is evaluated; whether both long-term and immediate changes in learning are to be assessed; and whether the measures of learning are closely related to the aims of learning through inquiry or reflect only the more general aims of science education.

Questions over-riding decisions on these matters relate to the purposes of the evaluation: who needs the information? who needs to be convinced of its value? and what decisions will be informed by the evaluation?

### ***Points from practice***

The paper takes just a few cases of evaluations of inquiry-based programs, and of some closely related innovations, which in their processes and outcomes have some bearing on the issues surrounding evaluation of impact on student learning. Key points arising from these cases concern:

- *Overcoming obstacles to implementation*  
Bringing about change in teaching may require both change in teachers and in school policies. Thus the support of the school principal is essential if innovations are to be properly implemented. Change is more likely when teachers have some part in the decisions about, and thus ownership of, the process of implementation in their classrooms. Classroom materials are too easily assimilated into existing practice. Teachers need firm commitment to change in order to withstand the influences that maintain the status quo.
- *The gradual process of implementation*  
Change from traditional to inquiry-based science teaching may require change in teachers' understanding of how children learn and of the nature of science. These changes are likely to be gradual; various researchers have identified a trajectory in the concerns of teachers when implementing reforms - from initial awareness, through personal and management concerns, to a focus on impact on students and on collaboration in and elaboration of the innovation. Most PD inputs may be too brief to take teachers far into this series of change; a well-founded study suggests that at least 80 hours is required.
- *The duration of implementation*  
Experience suggests that innovations should be operational in the classroom for at least one year, and preferably longer, before learning outcomes are assessed. During this time, information should be collected to determine the extent to which students' experiences are inquiry-based, that is, for example, the occurrence of collaborative learning, of use of process skills and of opportunity to raise and pursue their own

questions. The outcomes can then be interpreted in the light of the actual student experience.

- *Outcome measures*  
The choice of instrument for measuring student outcomes should reflect the purpose of the evaluation. Improved achievement on regular science tests are more likely to be convincing than on measures closely related to the goals of inquiry-based teaching. Both immediate and long-term changes in student learning should be evaluated, since some learning experiences take time to become internalised and applied by students in later learning.

### ***Implications***

To evaluate the effect of inquiry-based teaching on students' learning it is essential to ensure that the intended learning experiences are in place. The timescale or implementation needs to accommodate the fundamental shifts that are likely to be needed in teachers' thinking and beliefs about science and about education. Such changes cannot be made merely by placing new materials in teachers' hands or by providing two or three half-days of PD. Equally profound changes are required in students' thinking and ways of learning, which also will take time. Evaluation designers need to take account of the nature of these changes so that policy-makers and practitioners can have sound evidence on which to base decisions about the value of learning science through inquiry.

# Evaluating Inquiry-Based Science Developments

## ***Introduction***

This paper attempts an introduction to issues related to evaluation of inquiry-based teaching and learning in science, with particular attention to assessing any impact on students' achievement in science. Since many of these issues result from the nature of inquiry-based learning, the paper begins with a discussion of the nature of inquiry, of inquiry-based instruction and of the rationale for adopting it. The second part deals with issues of implementation and the difficulties of evaluating the effect of change in one area of students' experience in the context of very many other influences on their learning. The third section considers some issues of evaluation design. The fourth looks at a few cases of evaluation studies, which throw light on some of these issues, leading to the final section considering ways forward. The paper is written from a perspective that is overtly supportive of inquiry-based teaching and learning.

## ***1. The nature of inquiry and inquiry-based learning and teaching***

### ***1.1 Inquiry***

It is not difficult to find definitions of inquiry. The widely quoted description given in the National Science Education Standards (Box 1) captures the essence.

#### *Box 1*

Inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. Inquiry requires identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations. (NRC, 1996, p 23)

Scientists use inquiry to develop understanding of the natural and made world, because it leads to theories and ideas that explain observed events and phenomena. It also implies recognition that current theories give the best explanations we can at the present time, but that when evidence is found that conflicts with current theories, they have to be changed and alternatives found. Stephen Hawking explains this in Box 2.

When students are developing their understanding of the natural and made world around them, then, like scientists, they can use inquiry to arrive at ideas and theories that help them explain what they observe. Students also have to change their ideas as they encounter new and conflicting evidence. And, like scientists too, they do not begin from a clean slate, but from what they already know and the ideas they have already.

However, unlike scientists, students – particularly the young ones we are concerned with in elementary education – do not already have well developed skills of observing, collecting evidence, making predictions, testing possible explanations and interpreting findings. Thus a key aim of inquiry-based science education at the elementary level is to help students develop these skills (called inquiry, investigative or process skills). However it is a mistake to see skill development as the only focus of inquiry-based teaching and learning, for its purpose is also to develop understanding. To quote the National Standards again: ‘When engaging in inquiry students...identify assumptions, use critical and logical thinking, and consider alternative explanations. In this way students actively develop their understanding of science by combining scientific knowledge with reasoning and thinking skills.’ (NRC, 1996. p2)

### *Box 2*

Any physical theory is always provisional, in the sense that it is only a hypothesis: you can never prove it. No matter how many times the results of experiments agree with some theory, you can never be sure that the next time the result will not contradict the theory. On the other hand, you can disprove theory by finding even a single observation that disagrees with the predictions of the theory....Each time new experiments are observed to agree with the prediction the theory survives and our confidence in it is increased; but if ever a new observation is found to disagree, we have to abandon or modify the theory. At least that is supposed to happen, but you can always question the competence of the person who carried out the observation. (Hawking 1988, p. 10)

## **1.2 Inquiry-based learning in practice**

There are many ways in which inquiry can be practiced in schools: it is not a programme of study, not a scheme of work, or a curriculum model. What learning through inquiry means in practice can vary from the month-long series of activities, which develop into a year-long engagement of students in studying a vernal pond, described by Rankin (1999), to short investigations that might fit into the span of one or two lessons. Different modes of implementation ‘are not only inevitable but also desirable because they will paint a rich picture of meaningful learning in diverse situations. Multiple modes of inquiry teaching and learning will invite teachers to engage in participating in inquiry in ways that match their own beliefs and teaching styles’. (Keys and Bryan, 2001, 632).

What binds these different ways of implementation together, and makes them all recognisable as ‘inquiry-based teaching and learning’, are the actions that the teacher and students are engaged in. Box 3 summarises some key features of these actions.

*Box 3*

Teacher's actions	Students' actions
<ul style="list-style-type: none"> <li>• Providing experiences, materials, sources of information for students to use directly.</li> <li>• Showing the use of instruments or materials that students will need in their inquiry.</li> <li>• Asking open and person-centred questions to elicit present understandings and how students are explaining what they find.</li> <li>• Engaging students in suggesting how to test their ideas or answer their questions through investigation or finding evidence from secondary sources.</li> <li>• Where necessary, helping students with planning so that ideas are fairly tested.</li> <li>• Listening to students' ideas and taking them seriously.</li> <li>• Asking questions that encourage students to think about how to explain what they find.</li> <li>• Setting up opportunities for collaborative learning and dialogic talk.</li> <li>• Scaffolding alternative ideas that may explain the evidence from their investigation.</li> <li>• Gathering information, through observation, questioning and interaction, about students' developing skills and ideas.</li> </ul>	<ul style="list-style-type: none"> <li>• Engaging in exploration of materials, events, objects.</li> <li>• Working in collaborative groups, sharing ideas and constructing understanding together.</li> <li>• Raising questions and considering how answers may be found through investigation.</li> <li>• Proposing possible explanations of observations.</li> <li>• Suggesting how ideas behind possible explanations can be tested or questions answered through investigation/active inquiry.</li> <li>• Planning and carrying out investigations, making observations and measurements as appropriate, or using other ways of gathering evidence, to test ideas.</li> <li>• Keeping notes and recording results in suitable ways.</li> <li>• Relating results to the ideas being tested or question addressed; attempting to explain results.</li> <li>• Communicating what they have done; listening to and sharing ideas with others.</li> <li>• Reflecting on the process of the inquiry and on any change in ideas.</li> </ul>

The actions in Box 3 describe open or independent inquiry, where teachers allow students to develop their own questions and design their investigations. In practice such freedom is rarely possible and inquiry experienced by students is more likely to be 'structured' or 'guided', where students investigate questions identified by the teacher (Windschitl, 2001).

### **1.3 Why do we want children to learn in this way?**

The answer here has to be brief and restricted to three main and interconnected points. The first is that learning through inquiry accords with modern views of the psychology of learning, which sees learners having an active role in their learning. The 'behaviourist' view of learning (Skinner 1974), based on the idea that behaviours that are regularly rewarded will be reinforced and those that are repeatedly punished will disappear, places

the control of learning outside the learner. The learner is the recipient of ready-made understanding that has to be accepted and memorised. This is completely contrary to the view of learning as being constructed through the mental activity of learners, linking new and previous experience. Evidence of the active part that learners take in their learning comes not only from the work of learning psychologists but from those who study brain function (Bransford, Brown and Cocking, 1999). Piaget (eg 1929, 1955, 1956) led the way in showing that even young children strive to make sense of what they see around them. More recent work with babies also shows how they distinguish different features of object long before they can talk (Gopnick, Meltzoff and Kuhl, 1999). So their heads are by no means 'empty vessels' when children come to the classroom.

The second point concerns the nature of scientific knowledge. Much of science teaching in schools conveys a dated view of science, as being value-free, a succession of facts to be learned and reflecting 'truths' that somehow exist to be discovered by those clever enough to do so. By contrast, science is viewed today (eg Hawking, 1988) as a product of human thinking, yielding theories that are accepted as long as they fit the evidence available. The theories come from shared thinking among scientists and, as history shows, are subject to what is socially and culturally acceptable. (It took a hundred years for the ideas of Copernicus [that the earth and planets move round the sun] to be accepted even though this model fitted the observations far better than Ptolemy's model of the universe, mainly because the earth-centred model had been adopted by the Christian Church. Today, there are contemporary examples of theories based on cultural mores that are preferred to scientific theories in some societies.) So, if we regard learning science by students as a parallel of 'doing science' by scientists, then it is important that students discuss and compare their theories with others, combining their thinking about how to make sense of evidence. A learner is part of a group; contributes to the group thinking and gains much from it. These days the 'group' may not be physically in one place, but even when learning at a distance, one person is learning from and with others. The idea of 'social construction of ideas' arises from the work of Vygotsky (1978) who suggested that understanding exists in social interaction when learners as a group discuss evidence, attempt to give explanations and ask others to justify their explanations; individual learners build their understanding through this process.

Both of these points link to the third reason for inquiry-based instruction; that it enables us to take account of the ideas that students have already formed. The considerable volume of research revealing students' own ideas in science testifies to the way that children strive to make sense of the world around them, whether or not they are taught science. The ideas they use in explaining things make sense to them in terms of their (limited) experience and process skills. Although these ideas often seem strange and illogical to adults, it only takes a little reflection to see how students might come to hold them. These ideas reflect what the students have experienced and clearly indicate effort to make sense of their experiences. For example, the eight-year-old, who wrote about origin of the rust found on a nail in Box 4, was using experience of the observations of the form and occurrence of rust.

*Box 4*

There is a liquid in the nail which leaks out of the nail.  
 This forms big lumps as it leaks out.  
 This liquid only comes out when its wet.  
 There must be some sort of signal to tell it to leak.

*Box 5*

A thoughtful 12 year old girl said she found it difficult to believe some things her teacher said. For instance, the teacher explained the appearance of dew on grass on a cold morning in terms of water vapour in the air. The girl already had an explanation for this; her own idea that the coldness of the grass created the water. This idea also fitted her experience of the water drops on a bottle just after it was taken from the fridge. Although the teacher's idea did not make sense to her, she was not free to reject it; so it remained as an idea she knew about, and could recall if asked, but it was not part of her understanding.

The children's own ideas make sense to them and will be preferred to the 'scientific view' in the absence of convincing evidence that their own ideas are flawed and the alternative one makes better sense. In Box 5, the girl was not able to test either her own or the teacher's idea in a way that was convincing to her. There is a good deal of evidence that students hold their own non-scientific views along with the scientific ones when the latter are superimposed on the former rather than replacing them as a result of evidence gained through inquiry.

Scientific ideas not only fit evidence in a particular case, but also explain related phenomena. An important aim of science education is to enable students to develop more widely applicable ideas ('big' ideas) as these have greater explanatory power than 'small' ideas which explain specific instances. Inquiry, by starting from the ideas the children already have from previous experience, links these to new experience, so forming 'bigger' ideas. Thus by helping students to use and develop their ideas teachers are enabling them to understand a wide range of phenomena. Learning can't take place the other way round, by short-cutting the inquiry and teaching the 'big' ideas directly, for these are too abstract and too remote from students' experience to be meaningful. For example, students develop ideas about how some animals and plants in their back yard or park or stream depend on each other, from which they develop more general ideas about interdependence of living things. But if the big ideas are the starting point they may be grasped at no greater depth than slogans and this important relationship may never be truly understood.

Given that enabling students to learn through inquiry is a worthwhile aim, bringing about the typical actions of teachers and students indicated in Box 3 requires a considerable

shift in many classrooms. So we now turn to the ways in which this can be done and how we can find out how effective they are.

## **2. Implementing and evaluating inquiry-based teaching and learning**

### **2.1 Bringing about change in the classroom**

Ways of bringing about any change in the classroom from outside the school are a combination of one or more of the following:

- a) Direct change in students' experiences, through special experiences, or 'teacher-proof' materials, on-line or on paper;
- b) Change in teaching through provision of kits or programs comprising teachers' guide and other materials
- c) Change in teachers through provision of professional development, on-line or face-to-face.

a) Direct change in students' experiences can be brought about by special teaching or experiences such as summer science camps which are not part of regular school activities. There is evidence that inquiry-based experiences of this kind can affect learning (although most of the research has been done with high school students) and perhaps more importantly, attitudes to science, which are maintained over a considerable time (Gibson and Chase, 2002). However, such activities affect only a small number of students and for large-scale changes it is essential to influence the regular teaching. It is probably a myth that materials for young learners can be made that virtually by-pass the teacher and directly influence the students (teacher-proof). It may be possible to provide exercises in using algorithms in this way, but not to stimulate the kinds of student actions identified in Box 3. By their nature inquiry-based activities are difficult to anticipate and, while computer programs may be more interactive than paper-based materials, facilitation by the teacher is required to bring about the co-construction of understanding that is involved in learning science through inquiry. Consequently we will not discuss further this direct route to students.

b) Development of programs that set out procedures for teachers for inquiry-based activities (and that often provide materials for students as well) is a common way of bringing about change; attractive in holding out the possibility of reaching all schools willing to adopt the programs. But it is difficult to convey in writing only the reasons for the classroom interactions that are required if teachers are to develop students' understanding through inquiry, and there is much evidence that many users adopt procedures without commitment and adherence to the underlying values. For this reason some program producers restrict use to those who have also taken part in some in-service activities, though these are often minimal.

c) Using professional development (PD) as the main vehicle acknowledges that change in teaching is likely to need more than the provision of suitable classroom materials and

may require change in teachers' beliefs about science, about inquiry and about how students learn. Many PD programs (such as those provided by the Institute for Inquiry at the Exploratorium) attempt to model, through their procedures, the teaching actions that they wish teachers to adopt. They also give opportunity for teachers to experience inquiry at their own level, thus creating a firm base of understanding of the kind of learning that is intended for students. They reflect the six critical components of high quality PD identified by Supovitz and Turner (2000) in Box 6. These experiences take time, however, as does the reflection upon them and the discussion of implications for the teacher's role. Research, outlined later, indicates that 80 hours of PD are required for it to be effective. This is supported by experience of PD providers, who find change in practice increases in those who undertake more than one course.

### *Box 6*

High quality professional development must

- immerse participants in inquiry
- be both intensive and sustained
- engage teachers in concrete teaching tasks and be based on teachers' experience with students
- focus on subject-matter knowledge and deepen teachers' content skills
- be grounded in a common set of professional development standards
- be connected to other aspects of school change.

(from Supovitz and Turner 2000 p964/5)

## **2.2 Obstacles to evaluating the impact of programs and PD**

As noted above, improving the learning of students on a large scale has to be a two-stage process, in which inputs in the form of programs or PD are mediated by the teacher (Box 7). This introduces a number of variables that confound the evaluation of the impact and raises issues about the design and instruments used.

### ***Confounding variables***

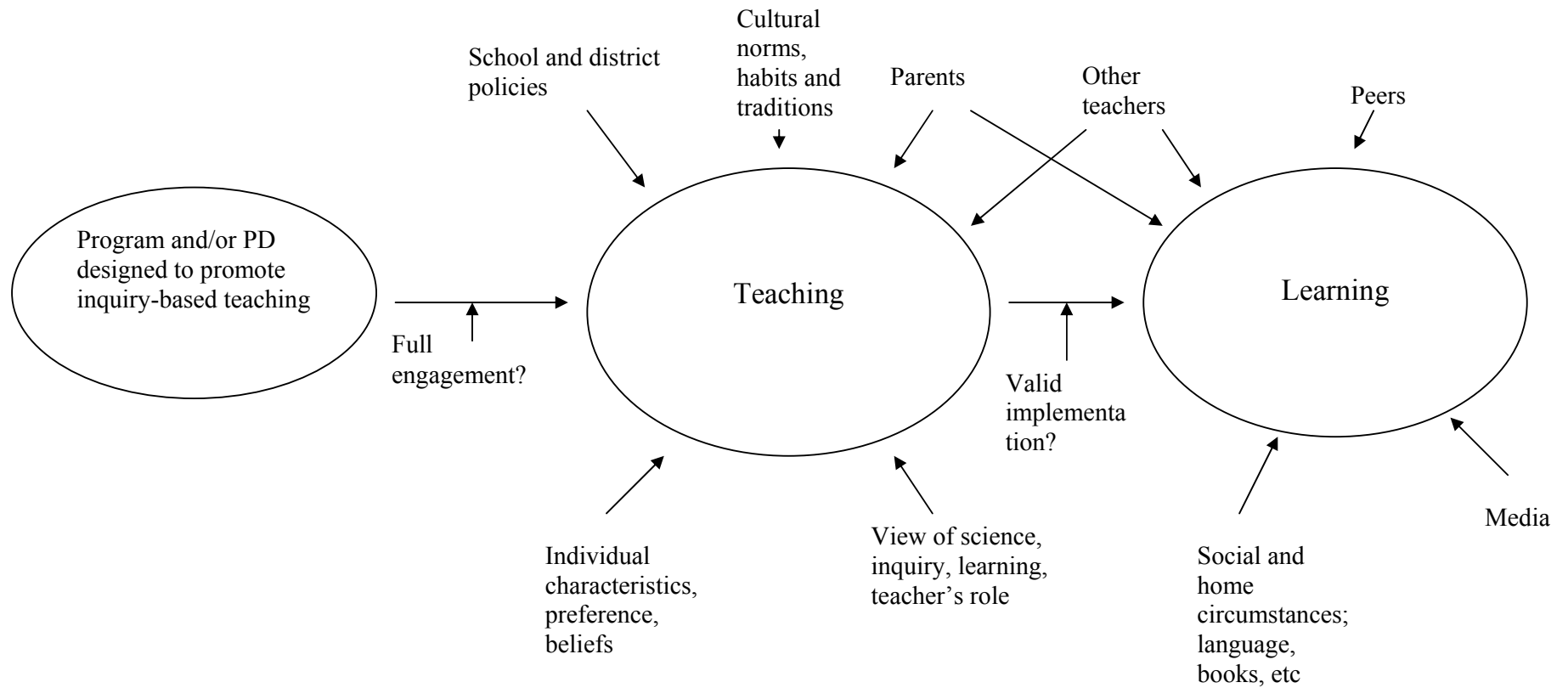
The question of how effective inquiry-based programs and PD are in bringing about changes in teaching and learning is notoriously difficult to answer. It only takes a little reflection to realise that there are very many influences on the processes at each step in the way from the creation of the program or PD design to the learner. Thus it may seem that any impact is likely to be so diluted as to be undetectable, even before one begins to consider how it might be detected. Box 7 sets out just some of the factors that might constrain implementation.

The link between the program or professional developers and the teacher is the initial threat to implementation; the teacher may not engage fully with, or may misunderstand, some of the intended messages. A teacher's ability to implement may be restricted by school policies, such as timetabling, or district policies, such as requirements to prepare students for certain tests. The practices of other teachers who teach the class may conflict with what is required for inquiry-based learning and may differ, for example, in the

extent to which students are expected to collaborate and discuss their work with each other. Thus students receive mixed messages about what actions are required for learning. Parents, too, may have certain expectations of what teachers should do, which differ from what a program requires. In addition there are the effects of the teachers' own background, experience and beliefs, which will transform, perhaps unconsciously, the messages of the program. Consequently the extent of implementation of the program as intended is problematic

Even if these other impacts on the teaching are not so great as to inhibit students having inquiry-based experiences, these will be only part of their total experiences. Other teachers, peers, parents and the media have direct or indirect influences on their learning.

## Box 7



### 3. Evaluation design issues

#### 3.1 Controlling variables

If the impact of inquiry-based experience is to be detected in the context of these other influences on students' learning, it is necessary to look for it in a situation where inquiry experience is the only difference between two groups whose learning is then compared. This means that the groups have to be similar in, for example:

- the content of science teaching,
- the characteristics and background of the students (gender, ethnicity, home language, etc),
- the teachers' background understanding of and attitude towards science and inquiry teaching,

to name but a few variables that are known to be associated with differences in learning. Controlling these variables to give equal 'experimental' and 'control' groups means being able to measure them, which itself is very difficult. Further, they are not independent of each other, so allocation to groups on the basis of one would cause inequality in another.

A way of avoiding this is to use a group as its own control, that is, to measure change over a period of time before the intervention (inquiry-based teaching) begins and then measure change over a similar period after intervention. Clearly there are problems with this because of the changes that take place anyway in the students over time.

Alternatively, in theory, students and teachers can be randomly allocated to experimental and control groups but this destroys normal student-student and student-teacher relationships and introduces an artificial situation from which results may not transfer to normal classrooms. It may also mean that teachers are reluctant to accept the role required or lack commitment to the innovation, leading to blurring of the distinction between the two groups.

Faced with this complexity and the seeming impossibility of detecting any change in learning that can be ascribed to an innovation, little attempt has been made by program and PD providers to assess change in students' achievements. Rather, they have tended to judge the validity of their programs in terms of the content and constructs represented and the responses of users of materials and participants in courses. Thus questions as to how to assess the outcomes have been avoided.

However, the issue is not only whether evaluation should focus on change in teaching *or* change in learning; both are needed. Knowing *that* students' learning has improved does not tell us about *how* this change was brought about. In the extreme case, an observed change may have had little directly to do with the program used or the PD experienced. Further, it can be argued that unless there is change in teaching, change in learning of the kind intended is unlikely. It is also necessary to know what kind of inquiry is being practised (open, guided or structured) in order to judge the impact of the initiative.

### 3.2 Selecting convincing measures of student outcomes

Despite the obstacles to detecting changes in learning that can be ascribed to the impact of inquiry-based programs, it remains the case that the ultimate judgement of the value of an intervention, once in place, must be whether students' learning is improved. Thus the question has to be addressed as to what measures would give evidence of 'success'? The main alternatives are:

- to assess outcomes that are specifically the goals of inquiry-based science education, such as use of process skills and ability to apply concepts;
- or, to use the regular methods of summative assessment (usually tests) which are likely to have a greater emphasis on factual knowledge and neglect process skills.

If only the first of these is used, then the results may not convince sceptics. It could legitimately be argued that, if comparison is made with a control group not having the inquiry-based experiences, no-one would expect anything other than better performance of the experimental group. Moreover, at present, there are not many measures available for assessing outcomes of inquiry-based learning that have the necessary validity and reliability.

If the second option is selected— using assessment or tests normally used — there is a chance that, at least in the short-term, no improvement in learning may be detected. However, given the claim of inquiry-based teaching to improve student learning in science generally, there is a strong argument that showing better achievement as measured by the accepted measures of learning is the most convincing evidence.

In either case there is a question as to how much change in the experimental group from pre- to post- inquiry-based teaching would be regarded as sufficient to justify policy and practice change. Would slightly more learning of the experimental group be enough – or considerably more? Who is to judge? Relying on statistics for this decision is problematic since what is statistically significant is not necessarily educationally significant.

### 3.3 Timescale

A further key design question relates to the timescale of an evaluation: what timescale would give convincing evidence of real benefits to learning? The important distinction is between looking only for immediate improvement in learning or for advantage that may still be evident several years later. It can be argued that immediate changes are of value, whether or not they are sustained, since students would be operating at a higher level than they otherwise would have been and their learning would necessarily benefit from this.

There are several points in favour of a longer timescale before assessing learning gains. First is the time it takes for teachers to become fully conversant with the reform and to be committed to it (this is point we return to later). Second is that the differences between teaching science through inquiry and teaching it in other ways are rather subtle and so the 'treatment' needs to be experienced over a considerable time for there to be a chance of long-term impact on levels of attainment. Third is that at the start of implementation, teachers can feel uncertain about their role. The change

required can reduce their self-confidence (previously based on an traditional view of a teacher's role) and produce temporary "de-skilling" until they begin to feel skilled in a new role. This is supported by experience (see the example of CASE later), that inquiry-base experiences take time to show in performance and there may be a danger, as Ruiz-Primo *et al* (2002) suggest, that premature judgments might be made 'by not finding effects when, although small, they are there and the reform that produced them should have been nurtured.'

Ways forward in relation to some of these issues can be inferred from studies of classroom events and evidence of change in teachers' understanding. Therefore, in the next section examples are given of recent research that is particularly informative about the factors that influence change, even though it has gone no further than looking at change in teaching. We then consider some of the limited number of studies that have looked at student outcomes.

## **4. Examples of evaluation studies**

### **4.1 Studies of impact on teaching**

Supovitz and Turner's (2000) study in 1997 of the local systemic change initiative collected information from 24 projects across the United States. Questionnaires provided self-reported data about teachers' attitudes, beliefs and teaching practices. Principals also responded to questions about their support for the initiative. From the large number of responses (787 schools and 4903 teachers) the researchers were able to develop scales of indicators of inquiry-based teaching practices. Using a multilevel model they linked the amount of PD experienced (among other teacher variables) to the indicators of teachers' inquiry-based teaching practice and levels of investigative classroom culture. They found that 'it was only after approximately 80 hours of professional development that teachers reported using inquiry-based teaching practices significantly more frequently ... than the average teacher' (p973). Their conclusions were that there was 'a strong and significant relationship between professional development and teachers' practices and classroom cultures.'(p 975), but the duration of most PD was too short to be effective. Their study also drew attention to other factors influencing change. For example, teachers from schools with low SES students used more traditional teaching methods than those with students of higher SES backgrounds. Another key influence was the supportiveness of the school principal.

The research into the TERC/Lesley University on-line Masters program for elementary and middle school science teachers showed that PD for inquiry-based science can be delivered with equal effect on-line as face-to-face, given a carefully crafted on-line course designed to provide experience of learning through inquiry. Harlen and Altobello (2003) compared changes in teachers studying the initial course in the program (Try Science) with a course with the same objectives and content delivered face-to-face. About half of the course time was spent learning science content through inquiry and half in studying aspects of inquiry-based teaching. In all parts of the courses, participants worked in groups. On-line participants were formed into groups of six or seven, each member exchanging asynchronous messages with the others every week. On-line participants conducted investigations or analysis of

evidence of teaching from videos or case studies off-line and then went on-line to report their results or ideas to their group. The research collected information about the course experience, changes in understanding of the science content, change in understanding of inquiry and inquiry-based teaching, change in confidence in teaching science and application of strategies for inquiry-based teaching in their classrooms. Although the main purpose of the research was to compare changes associated with on-line and face-to-face study, the relevance of the work here is the extent to which changes were found in teachers in either course.

The main findings are summarised in Box 8. For understanding of inquiry and understanding of inquiry teaching, both sets of teachers reported a greater change than was observable from more objective measures, such as the definitions they were asked to provide in pre- and post-course questionnaires. Thus there was a perception of development in understanding, which did not translate into practice. It was striking, too, that there was greater success in developing teachers' understanding in science than in bringing about change in classroom practice (judged for all participants from the lesson plans and their accounts of putting these into practice and, for five teachers, from observation of lessons).

#### *Box 8*

##### *Impact of the inquiry-based course in both modes:*

- The *Try Science* course, when delivered both on-line and on-campus successfully involved participants in scientific investigation in which they regularly used science inquiry skills.
- For both sets of course participants, there were changes in their understanding of the science content of the course, but this was significantly greater for the on-line participants.
- There was little change in the participants' understanding of inquiry evident in their pre- and post-course definitions, but those in both courses considered that their understanding of inquiry in science had been increased.
- Most participants initially held well-founded views about inquiry teaching and there was little change in this during the course.
- The lesson plans produced by participants in both courses made provision for students to undertake hands-on activity and to make predictions, but fell short of expectations in respect of students applying concepts and investigating their own questions.

##### *Differences between online and on-campus PD experience*

- The main difference in experience between the on-line and on-campus was that the former were involved more frequently than the latter in reflecting on their learning and on the process of inquiry.
- On-line participants experienced, appreciated and commented upon their collaborative learning; they did not feel that they were working alone.
- The confidence that teachers expressed in their capacity to teach science through inquiry increased during the course, significantly more for the on-line than for the on-campus participants.
- The average time spent on the course by on-line participants was approximately 90 hours and for the face-to-face participants was about 66 hours (of which 36 hours were in class).

Based on Harlen and Altobello, 2003

Those who have studied change in teachers have identified various stages in the implementation of new approaches, beginning with recognition and adoption of change without much understanding and ending with full understanding and commitment (Rudduck and Kelly, 1976; Doubler, 1991). A widely validated sequence of stages in implementation was developed by Hall and Loucks (1977) based on the concerns that teachers have in implementing reforms. Based on this theory Hord *et al* (1987) developed the series of stages in Box 9.

*Box 9*

0	Awareness	Little concern about or involvement in the reform
1	Informational	General awareness of the reform
2	Personal	Uncertainty about the demands of the reform and personal adequacy to meet those demands
3	Management	Concern focused upon efficiency, organizing, managing demands
4	Consequence	Focused upon impact of reform on students
5	Collaboration	Coordination and collaboration with other in using the reform
6	Refocusing	Exploration of possibility of major changes and alternatives

Although there may not be a generally agreed upon descriptions for the precise stages, what is accepted is that teachers pass through a trajectory of implementation, which is thus not a single event but a gradual change in understanding and in practice. It may also be different for various aspects of what is required. For instance some teachers may more easily work through to stage four in relation to changing their own questioning practice, whilst being at an earlier stage in, say, enabling students to investigate their own questions.

In the case of the participants in the *Try Science* courses, it appeared that the teachers were only at the beginning stages of change. Participants understanding of and commitment to providing experiences for students, required for inquiry-based learning, was not great enough to modify their existing practices. An introductory course would not be expected to do this. Anecdotal evidence indicates that as teachers continue to study inquiry on-line through the whole masters program they were more able to make the radical change in their role that are required to implement inquiry-based learning in their classrooms.

#### **4.2 Studies of impact on students' learning**

The issue of how closely the assessment of learning should match the learning activities was addressed in the research by Ruiz-Primo *et al* (2002) into changes in fifth grade students studying units in the FOSS curriculum. The multiple level approach of these researchers was based on test performance as the indicator of the effect of reform. They argued that that changes, if any, would show first in relation to performance in classroom activities, then in transfer to closely related activities involving different content, then in more distantly connected tasks, and so on. On this basis they identified methods of assessing learning at five levels of distance from the classroom learning: immediate, close, proximal, distal and remote. They developed

measures of change at these levels for two fifth grade science units of the FOSS program (Box 10).

The researchers used the students' notebooks, collected at the end of the eight months of the study, as a source of data about the implementation of the unit as well as a measure of 'immediate' learning. The 'close' and 'proximal' assessments were given as pre- and post-tests. After the post-tests, the 'distal' assessment was administered.

The results showed that, averaged over the 20 classes involved, there was a significant difference between pre- and post-test results for both the close and proximal assessment. As expected, differences were greater for close than for proximal measures and there was a considerable range across classes and between the two units that were taught. Judged from the students' notebooks, the researchers noted that 'the demands of the tasks required by the teachers ... were generally low. Teacher tended to ask students to record the results of experiments (not to interpret results) or to copy definitions. These tasks by themselves can hardly help students improve their understanding.' (p383). As there were no pre-test scores for the distal measure, correlations between distal and the close and proximal post-test scores were used to the effect on the distal assessments. These correlations were significant and about the same magnitude as each other, contrary to expectations.

*Box 10*

***The measures of immediate, close, proximal, distal levels for the FOSS Variables Unit (Ruiz-Primo et al, 2002)***

Immediate: assessment of artefacts during the curriculum activities; in particular how well variables were identified and manipulated in swinging a pendulum.

Close: assessment was a modified pendulum activity, with variation in materials used and the way that the dependent variable was measured.

Proximal: assessment was based on asking students to explain what makes bottles float or sink, thus students had to use knowledge of manipulation of variables and how to interpret results in a quite different task.

Distal: a test developed for large-scale assessment based on the National Science Education Standards and administered by the California Systemic Initiative Assessment Collaborative, using 'trash' as the subject matter.

Remote: no specific information given except that it would be an assessment of general educational achievement.

Although the conclusions from this study were tentative, they indicate that the 'distance' from the classroom of the assessment used in evaluation does matter.

The Cognitive Acceleration through Science Education (CASE) projects in the UK are not essentially inquiry-based interventions, but their aim to develop thinking processes makes them highly relevant in the present context. There are three CASE projects all sharing the goal of developing students' levels of thinking. In the first project, the target age group was 11 and 12 year olds and the aim of the classroom intervention was to accelerate thinking from Piaget's concrete operational to formal

operational levels. Adey and Shayer (1990) created a set of activities designed to develop students' identification of variables and the use of relationships between variables to make predictions, to stimulate cognitive conflict, and to encourage metacognition and the explicit linking of strategies to contexts outside those in which they were developed.

Teachers using the materials during the evaluation trials were given special training. During the trials visits were made to check on the way in which the intervention was being operated. Experimental and control classes were tested before and after the two year trial using a Piagetian Reasoning Test developed by the researchers. The post-tests also included science achievement tests developed in collaboration with the teachers to serve as end of school year tests. There were also immediate post-tests and delayed post-tests, one-year after the intervention and then, after a further year the results of the national examinations (GCSE) in science taken by all 16 year olds were collected. Thus the measures would be described as 'distal' in the terms used by Ruiz-Primo *et al* (2002).

The results immediately after the two year trial showed difference only for the Piagetian Reasoning Test, where significant differences from the control group were found only for 12 year old boys. One year later these differences had disappeared, but difference on the science post-tests began to appear. The effect on achievement in science appeared to be strengthening; the effect was even greater in the later GCSE results. The experimental group was significantly ahead in the science examinations, but there were also gains in mathematics and English, a finding possibly explained by the examination requiring analysis of characters and comprehension of texts, which could well have been influenced by the intervention activities. The results appear to show that general thinking skills can be taught. But they also show that the impact may well be delayed.

The success of CASE at the boundary between concrete and formal thinking encouraged the development of work at the lower age where, according to Piaget, students move from pre-operational to concrete operational thinking. Thus CASE was created for five to seven-year-olds, in a project called Let's Think. Here the activities have been developed around the Piagetian schema of concrete operations that is, seriation, classification, points of view, causality, rules and time sequencing. (Adey *et al*, 2002). Results of the year-long trials showed that students using the activities made significantly greater gains on a spatial perception tests and a conservation test than matched control groups. Since conservation is not part of the program, this shows transfer and suggests that these activities can raise students' general ability to think (Robertson, 2004).

As expected there was variation across classes and investigation of the causes of this add to understanding of the influences on implementation. An important and expected factor was the support of the school principal and senior management team. The pressure on schools in England to meet National Curriculum requirements can restrict teachers' freedom to give activities of this kind the time they require for proper implementation and to use them at the time of day when children are focused. The researchers also found that the activities were more effective when students work in groups of mixed gender, ethnicity, ability and personality. Students also made more

cognitive gains in classes where teachers attended to the issues surrounding successful group work, while keeping the criteria of ‘success’ clearly as ‘helping each other to think better’.

A third CASE project, using materials developed for eight and nine-year-olds is in progress. This will be able to use the National Tests given at age 11, which include science tests, to evaluate the impact on children’s learning.

In another English project (Black *et al*, 2003), where PD was the only input, National Tests in science were used as a measure of effectiveness. This project focused on developing teachers’ practice in formative assessment. Although not directly focused on inquiry, formative assessment requires very similar actions of teachers, in particular to find out and use information about students’ ideas and skills, share learning goals with students, encourage students to take responsibility for their learning and reflect on how they learn so that they can assess their own work. This project involved PD input over a period of two years, during which the main force for change was collaboration within groups of teachers rather than input from the researchers, who saw their role as responding to the teachers’ needs. The 19 teachers involved decided on the particular aspects of their practice that they wished to change in order to implement formative assessment, the main ones being questioning, feedback through marking, peer- and self-assessment, and the formative use of summative tests. Between PD meetings, the researcher visited classrooms to observe lessons and give feedback to the teachers on these observations. After two years of collaborative working the results for each class were compared with an equivalent class in the same school. The differences found could be expressed in terms of national curriculum levels (one level is roughly equivalent to two years): just under half a level for students in the age 11 tests and just over half a level for students in the age 13 tests. These are substantial gains.

Again, this project found variations across classes and identified factors that inhibited the implementation of practices considered desirable. These included a homework policy that requires grades on all work, whereas the project provided evidence that giving comments only increases levels of achievement (Butler, 1987). Another example was target-setting that requires frequent review and inhibits change in approach to learning. Thus it was important for the school management to be prepared to relax such requirements to give the teacher a chance to try out innovation.

## **5. Ways forward**

These cases, although only examples, throw light on the issues raised earlier in section 2. This section begins with a summary of these points, followed by some implications for future evaluations.

### **5.1 Research on change in teachers**

Key points are:

- Change in teaching may require both change in teachers and in school policies. So the support of school principals and senior management is essential for innovations to be properly implemented and so be validly evaluated.

- Change is more likely when teachers have some ownership of the process of implementation in their classrooms.
- Innovations should be in the classroom for at least one year, and preferably longer, before students' learning outcomes are assessed.
- Most PD is of too short a duration to have an enduring effect.
- Evaluation methods that use only self-report of teachers may suggest impact on teachers that is not necessarily translated into changes in practice.

These points suggest that an overarching concern should be to identify the point reached by teachers in their understanding of and implementation of the inquiry-based teaching. As trained professionals, teachers have established views and practices in relation to science teaching based on their understanding of how students learn and the nature of science (Harlen and Osborne, 1985). Making changes is a gradual process, as we have noted, with variation from one individual to another and within one individual in the degree to which certain aspects of the reform are understood and internalised. Thus, when the impact of reform on students is being investigated, it is crucial to determine the point reached by teachers in the trajectory of implementation, since this is likely to influence the experience of the students.

## 5.2 Research on change in learners

The main points are:

- Process information from classrooms should focus on the occurrence of collaborative learning, and of students using inquiry skills, pursuing their own questions and using evidence.
- Product information should enable comparison of observed change in attainment with what would have taken place over the same period of time without the intervention of inquiry-based science teaching.
- The choice of instruments for assessing learning outcomes may determine whether or not changes are found; the 'distance' of transfer is a factor to consider.
- Some experiences take time to impact on student's thinking and for them to apply this to thinking in further learning.
- Measures of outcomes that are convincing to the users of the evaluation should be applied; assessment of goals that could only be achieved by those experiencing the innovation are useful for research but not for policy-making.
- There need to be ways of communicating the size of the effect on learning in terms familiar to the audience.

Together these points underline the decisions in research design that, in addition to the intervention, influence evaluation outcomes. In making the decisions, the purpose of the evaluation has to be borne in mind, so that the information is relevant to intended users. They also further draw attention to the need to allow time for changes in teachers to work through to change in students' classroom experiences and then into changes in students' skills and ideas. The more worthwhile and enduring learning benefits are likely to be the ones that take more time to become evidence in assessments of learning outcomes.

### 5.3 Implications

Assuming that the small scale trials during development of PD programs or classroom materials have established that these initiatives are capable of bringing about learning through inquiry, the concern is with evaluation of their implementation in the context of regular schools and classes. As the glimpses of particular evaluations have shown, there are many factors that interfere in establishing the extent to which student outcomes can be detected. We can consider them in two main groups: those relating to the ‘independent variable’ i.e. the extent of students’ experience of inquiry; and those relating to the measurement of the dependent variable i.e. the impact on students.

In relation to the independent variable, existing practice suggests that serious attention needs to be given to ensuring that we know what any outcomes result from. This means gathering information about teachers’ understanding of the reform, classroom processes and student outcomes. Too many evaluation studies have reported (as, for example, Ruiz-Primo *et al*, 2002) that classroom events showed few of the practices associated with inquiry learning. In such cases, the results do not provide the basis for a valid judgement of the value of the program. The strong evidence of the influence of the principal indicates that there needs to be involvement of senior management in decisions about initiatives in order to ensure more effective implementation.

When there is evidence that the independent variable is in place, it needs to be in operation for at least a year before outcome measures are applied. This timeframe acknowledges all the other impacts on student outcomes, and that inquiry is aimed at changes in ways of learning that take time to establish.

Turning to the dependent variable, there are important decisions to be taken about the nature of the measures that are used, be they to assess attitudes, achievement, choice of further study, etc. As discussed earlier, the measures have to convince those who need to be informed about and/or will make decisions on the basis of the evaluation. They also need to be applied over a period of time so that questions can be answered about short- and long-term changes in learning.

The design of the evaluation has to allow observed changes to be compared with what would have taken place without the intervention. But the magnitude of any difference will need to be expressed in meaningful terms – percentage differences are unlikely to be helpful and statistical significance does not imply educational significance, particularly where numbers involved are large. Black and Wiliam (1998), for instance, found it helped to impress policy-makers by pointing out the potential gains from implementing formative assessment would, if applied across the country, have raised England from the middle of the 41 countries involved in the TIMSS study to being one of the top five.

Finally evaluation designs should take account of the profound changes that are required in teachers’ and in students’ thinking. Only if time is allowed for these to become embedded in practice will evaluations be able to provide policy-makers and practitioners with valid information about the impact on students.

## References

- Adey, P., Robertson, A. and Venville, G (2002) Effects of a cognitive acceleration programme on year 1 pupils. *British Journal of Educational Psychology*, 72, 1 – 25
- Adey, P. and Shayer, M (1990) Accelerating the development of formal thinking in middle and high schools students. *Journal of Research in Science Teaching*, 27 (3) 267 – 285
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003) *Assessment for Learning. Putting it into practice*. Maidenhead: Open University Press.
- Black, P, and Wiliam, D. (1998) *Inside the Black Box*. School of Education, King's College, London
- Bransford, J., Brown, A.L. and Cocking, R.R (Eds) *How People Learn*. Washington DC: National Academy Press
- Butler, R. (1998) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology* 58, 1 – 14
- Doubler, S. (1991) Change in Elementary School Teachers' Practice in the United States. Unpublished PhD thesis, University of Liverpool.
- Gibson, H. L. and Chase, C. (2002) Longitudinal impact of an inquiry-based science program on middle school students' attitudes towards science. *Science Education* 86 693 - 705
- Gopnik, A., Meltzoff, A. N. and Kuhl, P. K. (1999) *The Scientist in the Crib*. New York: William Morrow
- Hall, G.E. and Loucks, S.F. (1977) A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*, 14 (3) 263 – 76)
- Harlen, W. and Altobello, C. ( 2003) *An Investigation of Try Science Studied On-line and Face-to-Face*. Cambridge, MA: TERC
- Harlen, W. and Osborne, R. (1985) A model for learning and teaching applied to primary science. *Journal of Curriculum Studies*, 17 (2), 133-146
- Hawking, S. W. (1988) *A Brief History of Time*. London: Bantam Press.
- Hord, S. M., Huling-Austin, L., Hall, G. E. and Rutherford, W. (1987) *Taking Charge of Change*. Alexandria, VA: ASCD
- Keys, C. W. and Bryan, L. A. (2001) Co-constructing inquiry-based science with teachers: essential research for lasting reform. *Journal of Research in Science Teaching* 38 (6) 631-645
- National Research Council (1996) *The National Science Education Standards*. Washington DC: National Academy Press
- Piaget, J. (1929) *The Child's Construction of the World*. London: Routledge and Kegan Paul
- Piaget, J. (1955) *The Child's Construction of the Reality*. London: Routledge and Kegan Paul
- Piaget, J. (1956) *The Child's Construction of Space*. London: Routledge and Kegan Paul
- Rankin, L. (1999) Lessons in a pond: a year-long inquiry investigation. *Foundations 2* Arlington, VA: NSF 71 - 78

- Robertson, A. (2004) Let's Think: Two years on. *Primary Science Review*, 82, 4-7
- Rudduck, J. and Kelly, P (1976) *The Dissemination of Curriculum Development: Current Trends*. Slough: The National Foundation for Educational Research
- Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L. and Klein, S. (2002). On the evaluation of Systemic Science Education Reform: searching for instructional sensitivity. *Journal of Research in Science Teaching* 39 (5) 369 393
- Skinner, B.F. (1974) *About behaviourism*. New York: Alfred A. Knopf
- Supovitz, J. A. and Turner, H. M. (2000) The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37 (9) 963 – 980.
- Vygotsky, L.S. (1978) *Mind in Society*. Cambridge, MA: Harvard University Press
- Windschitl, M. (2003) Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Teacher Education*, 87, 112- 143