

As published in the *British Educational Research Journal*, 30, 3, 2004, pp 359-378.

Reasoning as a scientist: ways of helping children to use language to learn science

Neil Mercer*, Lyn Dawes**, Rupert Wegerif*, Claire Sams*

(Open University*) (De Montfort University, Bedford**)

Abstract

Sociocultural researchers have claimed that students' learning of science is a discursive process, with scientific concepts and ways of reasoning being learned through engagement in practical enquiry and social interaction as well as individualised activity. It is also often claimed that interacting with partners while carrying out scientific investigations is beneficial to students' learning and the development of their understanding. The research we describe investigated the validity of these claims and explored their educational implications. An experimental teaching programme was designed to enable children in British primary schools to talk and reason together and to apply these skills in their study of science. The results obtained indicate that (a) children can be enabled to use talk more effectively as a tool for reasoning and (b) talk-based activities can have a useful function in scaffolding the development of reasoning and scientific understanding. The implications of the findings for educational policy and practice are discussed.

Corresponding author: Professor Neil Mercer, Educational Dialogue Research Unit, Faculty of Education and Language Studies, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK. Email: n.m.mercer@open.ac.uk

Introduction

Educational researchers who adopt a sociocultural perspective have commonly depicted science education as a discursive process, whereby novices (students) are inducted into a way of representing and understanding phenomena (using language and other representational means) by those more expert in the field. Thus Lemke (1990) proposed that science education should enable students to become ‘fluent speakers of science’, while Leach and Scott suggest that students be helped by their science teacher to ‘make sense of the talk which surrounds them, and in doing so, relate it to their existing ideas and ways of thinking’ (Leach and Scott, 1995, p.44). Contemporary sociocultural theorists (e.g. Wertsch, 1991; Wells, 1999; Daniels, 2001) follow Vygotsky (e.g. 1978) in emphasising the importance of language use and social interaction within communities for the development of educated ways of making sense of the world, such as those associated with science. As Vygotsky put it, intermental (social) activity will promote intramental (individual) intellectual development. This claim, having an obvious plausibility, has been widely accepted. However, other than our own findings presented in an earlier issue of this journal (Mercer, Wegerif and Dawes, 1999), any empirical evidence offered for its validity has been, at best, indirect. Our earlier research showed that the induction of children into an explicit, collaborative style of reasoning which we call *Exploratory Talk* led to gains in children’s individual scores on the Raven’s Progressive Matrices test of non-verbal reasoning. These gains, first demonstrated for children in Year 5 in British primary schools, were subsequently replicated in other year groups and in primary schools in Mexico (Rojas-Drummond, Mercer & Dabrowski, 2001). However, our earlier studies did not investigate whether this induction was beneficial to children’s

learning in science. To the best of our knowledge, no direct relation has been demonstrated between encouraging students to engage in certain ways of using spoken language and their improved understanding or attainment in science. One group of studies has shown that discussion can contribute to the development of conceptual understanding in science (Howe, Tolmie, Duchak-Tanner & Rattray, 2000), but that research did not encourage the pragmatic use of particular forms of dialogue. Nevertheless, Howe *et al.*'s work is interesting and very relevant here. Their findings provide support for the value of discussion and investigation by children without the authoritative presence of a teacher, while also showing that expert involvement can have a crucial and beneficial influence for guiding children's activity in productive directions. On the basis of their experimental comparisons of different types of hypothesis testing task (with children aged 9-11), they specify some conditions for a kind of activity which effectively promotes the development of both conceptual and procedural knowledge. These are hypothesis-testing activities with a four-part structure, in which (a) pupils first debate their conceptual understanding and reach a consensus about the hypothesis to be investigated (with the pursuit of consensus a key requirement), (b) next subject their consensual positions to expert guidance (by a teacher) about how to pursue a practical controlled investigation of their hypothesis, (c) perform the investigation and (d) discuss the outcomes together to draw conclusions. These conclusions have strong and direct relevance to the design of our own study which, like those of Howe *et al.*, involved the use of computer-based group activities.

The aims and focus of the research

The research we describe in this paper was designed to test sociocultural claims about the significance of discursive induction for the study of science. The research also had the practical aim of improving children's ability to use language as a tool for reasoning and to improve their attainment in science and mathematics. (For reasons of space and clarity, this paper will only deal with science.) The study was based in primary schools in South East England and involved the implementation and evaluation of an experimental teaching programme for children in Year 5 (ages 9-10). One link between the learning of science and the use of language is the development of a specialised vocabulary for representing concepts and describing processes. In addition spoken language provides a familiar medium through which a child can describe their conceptions of phenomena in order that teachers assess a level of understanding (Ollerenshaw & Ritchie 1998). However, supporting the learning of vocabulary and descriptive skills was not the main focus of our research. Instead, our experimental programme aimed to develop children's use of language as a tool for reasoning together and to enable them to use the discussion skills they developed in carrying out science activities. This focus reflects the historical origins of the project in a series of earlier projects concerned with talk and reasoning in the classroom.

Contexts for teaching and learning science

There are two main contexts in which spoken language can be related to the learning of science in schools. The first is teacher-led interaction with pupils. A sociocultural account of cognitive development emphasises the guiding role of more knowledgeable members of communities in the development of children's knowledge

and understanding, and in their induction into the discourses associated with particular knowledge domains. The concept of ‘scaffolding’, as originally used by Wood, Bruner and Ross (1976) is relevant here, as are the concepts of ‘guided participation’ (Rogoff, 1990), ‘the guided construction of knowledge’ (Mercer, 1995) and ‘dialogic teaching’ (Alexander, 2003). For educational researchers, sociocultural theory highlights the role of teachers in helping children develop new ways of describing and conceptualising experience.

The second context is that of peer group interaction. Working in pairs or groups, children are involved in interactions which are more ‘symmetrical’ than those of teacher-pupil discourse and so have different kinds of opportunities for developing reasoned arguments, describing observed events, etc. In science education, such collaboration can be focused on practical investigations, which also have great potential value for helping children to relate their developing understanding of abstract ideas to the physical world. Computer-based science activities can offer similar opportunities in virtual environments. The research of Howe *et al.* (*op.cit.*), described earlier, has shown that under certain conditions computer-based activities for groups of children are effective in promoting the development of scientific understanding; but discussions amongst groups of young science students may not always be productive and useful. Observational research in British primary schools has shown that the talk which takes place when children are asked to work together is often uncooperative, off-task, inequitable and ultimately unproductive. (Galton & Willamson, 1992; Wegerif & Scrimshaw, 1997).

A possible explanation for the doubtful quality of much collaborative talk is that children do not bring to this task a clear conception what they are expected to do, or

what would constitute a good, effective discussion. This is not surprising, as many children may rarely encounter examples of such discussion in their lives out of school – and teachers rarely make their own expectations or criteria for effective discussion explicit to children (Mercer, 1995). Children are rarely offered guidance or training in how to communicate effectively in groups. Even when the aim of talk is made explicit – ‘Talk together to decide’; ‘Discuss this in your groups’ – there may be no real understanding of *how to* talk together or for what purpose. Children cannot be expected to bring to a task a well-developed capacity for reasoned dialogue. This is especially true for the kinds of discursive skills which are important for learning and practising science: describing observations clearly, reasoning about causes and effects, posing precise questions, formulating hypotheses, critically examining competing explanations, summarising results, and so on. On this basis, we began this research with the hypothesis that children studying science would benefit from teacher guidance of two main kinds. First and most obviously, they need to be helped to gain relevant knowledge of natural phenomena, investigative procedures, scientific concepts and terms – the content of science. Teachers commonly expect to provide this kind of guidance. Secondly, they need to be helped to learn how to use language to enquire, reason, and consider information together, to share and negotiate their ideas, and to make joint decisions. This kind of guidance is not usually offered. We therefore designed a teaching programme which would enable teachers to integrate these two kinds of guidance.

Method

The research involved the design, implementation and evaluation of a programme of

lessons. We will describe the intervention programme, the design of the study, how data was gathered and the methods used to analyse it.

The intervention programme

As mentioned in Section 1, our research has shown that language skills associated with improved reasoning can be effectively taught and learned. That earlier research involved the creation and evaluation of talk-based classroom activities. An intervention programme was designed for the present study which built directly upon that earlier research. A key feature of this ‘Thinking Together’ programme was the systematic integration of teacher-led interaction and group-based interaction. The main aims of the programme were:

- (i) to raise children’s awareness of the use of spoken language as a means for thinking together;
- (ii) to enable children to develop their abilities to use language as a tool for thinking, both collectively and alone;
- (iii) to enable children apply the tool of language effectively to their study of the science and maths curriculum.

More specifically, the programme was intended to ensure that children became able to carry out the kind of discussion we call *Exploratory Talk*. This is talk in which:

- All relevant information is shared
- All members of the group are invited to contribute to the discussion

- Opinions and ideas are respected and considered
- Everyone is asked to make their reasons clear
- Challenges and alternatives are made explicit and are negotiated
- The group seeks to reach agreement before taking a decision or acting.

Barnes and Todd (1977, p. 126) in their classic study of group work describe this sort of talk, which is not commonly heard, but which has great educational value:

‘It is a collective relationship that we observed in our small group discussions.

Members were free to shift the topic, to try out new formulations and to explore alternatives, since none of the questions asked concealed positional claims to impose a frame on the discussion – to guide its direction or to judge the relevance of answers. The members of our groups cast their bread upon the waters. They were each other’s resources and most of their utterances were contributions to thinking.’

Each teacher was provided with 12 detailed lesson plans. These lessons involved a teacher-led introduction, a group discussion activity, and a final ‘sharing’ plenary session. The aims of the lessons were to do with the teaching and learning of explicit talk skills such as critical questioning, sharing information, or negotiating a decision. These closely related to the speaking and listening component of the National Curriculum for English. The first five lessons were aimed at raising children’s awareness of how talk could be used for working together and establishing in each class a set of ‘ground rules’ for discussion which would generate talk of an ‘exploratory’ kind. The further seven lessons encouraged children to apply their developing discussion skills to the study of the science and maths curriculum for Year

5. Each lesson applied a specific talk skill and targeted a specific concept in science.

Some lessons involved computer-based activities. One of the software items chosen was Granada's *Science Explorer II*. This provides sets of problems and simulated experimental environments and was already in use in both target and control schools. The topics of 'light' and 'sound' (which are included in the National Curriculum scheme of work for Year 5) were the focus for these activities. For example, in one Science Explorer activity children were given the problem of sound-proofing a room against the singing of an opera-singer neighbour using such materials as wood, metal and cork. In relation to the study of light, the activity requires them to make a light-proof screen from a set of materials including tissue paper, writing paper and card. The task is to select a type of (virtual) material and predict the number of sheets predict which will block out light. The group then test their prediction in the simulated environment. The content and level of these problems are suited well to the Year 5 Science curriculum in English schools, as prescribed for the classes involved in our study. Although we gathered data from activities based on several types of software, in this paper our illustrative examples of talk come from the use of *Science Explorer*. In the context of the intervention, children were prompted to discuss their predictions and findings while engaged in the activity. (The particular use of Science Explorer in the Thinking Together activities is discussed in more detail in Wegerif and Mercer, *in press*.)

A necessary condition for the implementation of the intervention programme was that teachers would effectively model and guide the development of children's language skills. Accordingly, each participating teacher received training in the Thinking Together approach, based on videotaped examples and activities derived from earlier

related projects. Once the intervention was underway, this training was reinforced through regular visits to schools by one of the research team, which on some occasions included demonstrating activities and teaching techniques. Researchers tried to ensure that all the lessons in the programme were carried out in each school. This was achieved for the initial 5 lessons, but for a variety of practical reasons some teachers were unable to fit all the later lessons into their class's activities.

Design of the study

The effects of this programme on talk, reasoning and learning were studied through observation and formal assessment of children in experimental classes, with pre- and post-intervention comparisons being made with children in matched control classes in other local schools with similar catchments. Seven classes of children in Year 5 (aged 9-10) in primary schools in Milton Keynes participated actively by following the Thinking Together intervention programme. These classes were designated 'target classes'. There were 196 children in the target classes at the beginning of the study, but with children arriving and departing from classes during the intervention period, 109 of the original children completed the programme. (Milton Keynes has a high rate of population movement, within the city as well as in and out of it.) A matched set of 'control' classes in other, similar local schools were identified. These consisted of 210 children at the beginning of the study, with 121 of those still being in the control classes by the time the last data collection was made. Control classes did not participate in the Thinking Together programme, but followed the same prescribed English and science National Curriculum and also used the Science Explorer software, but not necessarily as a basis for regular group activities.

The teachers of the target classes were given training relating to the first part of the project on an initial training day and the first 5 lessons were completed by the end of the first term (which also saw the completion of the pre-intervention observations and assessment). Teachers of control classes did not receive training but were aware that the results of the research would be made available to them on its completion. We have no reason to believe there was any relevant exchange of ideas between target and control schools during the period of study. The complete set of the lessons (including those concerned with mathematics, which are not discussed here) were implemented in the target schools over the Autumn and Spring school terms, with the total duration of the intervention (i.e. from the pre-testing to the post-testing) being 23 weeks.

Data collection

The data gathered consisted of:

- (i) pre- and post-intervention video recordings of a ‘focal group’ in each target and control class carrying out computer-based activities;
- (ii) video recordings of other groups of children in target schools engaged in joint activities during Thinking Together lessons;
- (iii) video recordings of teacher-led whole class sessions during Thinking Together lessons;
- (iv) audio-recordings of interviews with teachers and children;
- (v) children’s scores on the Raven’s Progressive Matrices test of non-verbal reasoning;

- (vi) children's scores on tests of knowledge and understanding in science (based on a set of assessment tasks for Year 5, known as 'optional SATs', which are made available to schools in England and Wales by the Qualifications and Curriculum Authority).

Pre- and post-intervention data was gathered for all of these in target schools, with pre- and post-intervention data for (i), (iv) and (v) also gathered in the control schools.

Interviews and video recorded interactions were transcribed. The data we gathered was intended to inform us about:

- (a) The teacher's role in scaffolding children's use of language and their learning of science
- (b) the nature and content of children's discussions;
- (c) the outcomes of our intervention on the quality of children's discussion;
- (d) the outcomes of our intervention on children's understanding of science

To study the teacher's activities in guiding children's use of talk and learning, we video-recorded all the target class teachers carrying out at least two of the prescribed lessons. (For reasons of space, we cannot discuss the analysis of the teacher's role in this paper: see Rojas-Drummond and Mercer, *in press*, for more on that topic). In order to study the effects of the Thinking Together programme on curriculum-related learning, we compared target and control children's understanding of National Curriculum science topics that they had studied during the period when the intervention took place. We describe this assessment in more detail below.

Methods of analysis

Qualitative analysis of children's talk in groups

Using the video-recorded data, qualitative and quantitative methods of discourse analysis were used to investigate changes in the quality of children's talk and collective reasoning. The methodology for making this kind of comparison, described fully in Wegerif and Mercer (1997), combines a detailed qualitative analysis of language used by each group of children in specific episodes of joint activity with a quantitative computer-based analysis of the whole corpus of recorded group talk.

These methods were used to make two kinds of comparisons:

- (a) between the talk of children in control classes and target classes doing curriculum-based activities at a stage when the 'ground rules' for talk had been established in target classes; and
- (b) between the pre-intervention and post-intervention talk of children in target classes while involved in similar activities.

We hoped in this way to feel confident that any differences observed in the quality of children's talk (in ways that related to our hypotheses) were attributable to the intervention.

The recorded data of children talking as they carried out activities in the Thinking Together programme was subject to joint qualitative analysis by the research team. The main aim of the analysis was to gauge the extent to which the discussions of children in the target classes came to resemble Exploratory Talk. However, the analysis had more subtle aspects than this may seem to suggest. We were not interested in simply identifying stretches of talk as 'exploratory' and coding them

accordingly. Indeed, our view is that the nature of language – in which any one grammatical form can be used to fulfil a range of pragmatic functions – renders any coding scheme of dubious value if used separately from a more contextually sensitive, ethnographic type of analysis. Our aim is to consider the extent to which children are using language appropriately and effectively as a tool for thinking together. In making that analysis, the definition of Exploratory Talk serves as an ‘ideal type’ – a typification of reasoning embodied in talk. The features of Exploratory Talk, as described earlier, are therefore used as a point of reference for the consideration of the quality of the talk of each group. (See Wegerif & Mercer, 1997 for more detail)

One focus of the qualitative analysis was on changes in children’s use of talk while engaged in science investigations. This involved the examination of the pre- and post-intervention talk of children in target classes as they engaged in computer-based science investigations using Science Explorer II and related to the Year 5 curriculum. The aim of this analysis was to determine if children’s joint engagement with the task changed in the predicted ways following the intervention: that is, if they used more talk of an ‘exploratory’ kind and applied it productively to a scientific task.

This detailed qualitative discourse analysis of talk by the researchers was used in combination with other methods to determine whether children in target classes began to use more Exploratory Talk and apply it appropriately through participating in the project. One additional method was a ‘blind assessment’. The purpose of this was to avoid biased judgements arising from the research team’s expectations about the outcomes of the intervention. To enable this, one group of children in each target and control class were observed and recorded using the same piece of software in the same activity at the beginning and again at the end of the programme. It was

originally intended that the activity to be used for this ‘blind’ pre/post comparison would also be based on a *Science Explorer* activity. However, on reflection it was agreed that the use of one of the activities devised for the project programme would not constitute a fair test, as the target children could be influenced positively by their greater familiarity with discussions based on the Science Explorer environment. A decision was therefore taken to base comparisons of children’s talk on recordings of focal groups carrying out a ‘moral dilemma’ activity based on software called *Kate’s Choice* (designed by a member of the research team and tested in earlier, related research) since this would be novel to all the children involved.

A set of 12 tapes and transcripts of the pre- and post-intervention discussions of 6 focal groups in target classes as they carried out one computer-based activity were modified to eliminate all clues as to location and time of recording. Two Open University researchers not involved in the project, but who were familiar with the analytic method and with classroom interaction, studied this data. They were asked to jointly (a) identify episodes of talk which had the characteristics of Exploratory Talk; (b) note the extent of these within the recorded event; and (c) use their findings to decide which discussions were ‘pre-intervention’ and which were ‘post-intervention’.

Quantitative analysis of differences in talk

A quantitative analysis of pre/post differences in the talk of target groups was made using the same *Kate’s Choice* data. As in our previous research, one aspect of this analysis focused on the relative extent to which children used ‘indicator words’ which qualitative analysis in earlier research had shown were associated with reasoning. The indicator words selected were ‘because’, ‘if’, ‘I think’, ‘would’ and ‘could’. Previous research had also shown that the more children explain and justify their views, the

longer their utterances will tend to be. We therefore compared the relative mean length of pre/post-intervention utterances in the focal groups when involved in the same type of activity.

Evaluating knowledge and understanding of the science curriculum

Improvements in subject-related knowledge and understanding were assessed in the following ways. First, as mentioned earlier, we used problems taken from the ‘optional SATS’ tests for Key Stage 2 provided by the QCA. These covered areas of the curriculum studied by children in both target and control classes. Secondly a ‘concept map’ activity was devised to assess individual understanding of scientific terms and their relationships (as used in some earlier research on science education (Russell and Watt, 1990) and computer use in schools (Harrison *et al.* 2002)). This activity was based on the topics of ‘food’ and ‘healthy eating’ which were studied by both target and control classes. Children were provided with a list of key words and were asked to make a concept map by linking words and annotating these links. Children’s responses were scored for the number of links and labels in their concept map which related appropriately to the relevant curriculum knowledge. For example a link between the word HEALTHY and the word FRUIT would score one point and an appropriate annotation on that link such as ‘VITAMINS’ or ‘FIBRE’ would score another point. The same activity was given to target and control classes at the beginning and at the end of the school year. Additional evidence came from teachers’ assessments of children’s progress made over the same period.

Changes in the quality of collective and individual reasoning

As mentioned earlier, our previous research had indicated that teaching children to

use Exploratory Talk helped them to solve the non-verbal reasoning problems of the Raven's Progressive Matrices test more effectively (Mercer, Wegerif & Dawes, 1999; Wegerif, Mercer & Dawes, 1999). We compared pre- and post-intervention performances of target and control classes working on the problems of the Raven's test in groups and as individuals. We have used this finding to argue in support of a Vygotskian, sociocultural account of the relationship between language use and intellectual development. That is, we believe that by engaging in guided, spoken reasoning with their peers, children can assimilate a way of thinking which helps them to reason better when working alone. The Raven's test is an appropriate measure for supporting this claim, as its scores correlate highly with other tests of reasoning and with measures of academic achievement (Raven, Court & Raven, 1995). However, in the earlier research, small sample size meant that although findings for individuals were statistically significant, findings for group scores did not achieve statistical significance (though they were in the predicted direction). This represented a weakness in the evidence we offered for the link between intermental activity and intramental development. In the current project we therefore attempted the test of this sociocultural hypothesis again, with a larger sample of children. Again, the Raven's test was used to compare the pre- and post-intervention both individual and group performances of target and control classes.

Results

Our main hypothesis was that if teachers focused on the direct teaching of ways of using talk effectively for joint reasoning and on the development of children's awareness of the importance and value of this use of talk during science activities, this would raise children's achievement in the study of science. We also predicted

effects of the intervention on the quality of children's talk and interaction and on their non-verbal reasoning. Analysis of the data supported these predictions, as explained below.

Changes in the quality of talk and collaborative activity

Qualitative analysis of talk

We have selected brief but representative sections of the many hours of talk we recorded to illustrate the patterns we discerned in our analysis. The two examples below illustrate the kinds of differences between the talk of target and control groups that we observed. In the first example, a group in a control school are working on a Science Explorer investigation into the effectiveness of materials for providing soundproofing.

Transcript 1: Control school group: Keep it Quiet

Hannah *(reads from screen)* 'Keep it Quiet. Which material is the best insulation? Click 'measure' to take a sound reading. Does the pitch make a difference?'

Darryl No we don't want clothes. See what one it is then. *(Points to screen)*

Hannah No it's cloth.

Darryl Oh it's cloth.

Hannah Go down. This is better when Stephanie's in our group.

Darryl Metal?

Hannah Right try it.

Deborah Try what? That?

Hannah Try 'glass'

Darryl Yeah.

Deborah No one.

Hannah Now

Darryl (*interrupts*) Measure.

Hannah Now measure. Hold. (*Turns volume control dial below screen*)

Darryl Results, notes.

Hannah Results. We need to go on a different one now. Results.

Darryl Yeah, you need to go there so you can write everything down.

Hannah I'm not writing.

For comparison, the next transcript is of a group of children who have been involved for two terms in the Thinking Together programme. They are engaged in a Science Explorer activity about the effectiveness of materials for blocking out light.

Transcript 2: Target school group: Blocking out light

Ross OK. (*reads from screen*) 'Talk together about a plan to test all the different types of paper.'

Alana Dijek, how much did you think it would be for tissue paper?

- Dijek At least ten because tissue paper is thin. Tissue paper can wear out and you can see through, other people in the way, and light can shine in it.
- Alana OK. Thanks.
- Alana (*to Ross*) Why do you think it?
- Ross Because I tested it before!
- Alana No, Ross, what did you think? How much did you think? Tissue paper. How much tissue paper did you think it would be to block out the light?
- Ross At first I thought it would be five, but second -
- Alana Why did you think that?
- Ross Because when it was in the overhead projector you could see a little bit of it, but not all of it, so I thought it would be like, five to block out the light.
- Alana That's a good reason. I thought, I thought it would be between five and seven because, I thought it would be between five and seven because normally when you're at home if you lay it on top, with one sheet you can see through but if you lay on about five or six pieces on top you can't see through. So that's why I was thinking about five or six.

The talk of the group in Transcript 1 indicates that the children are not working co-operatively except at the most superficial level. They do not share knowledge, build on each other's suggestions, provide reasons for their proposals or seek joint agreement. The contributions are monosyllabic and minimal. It does not seem that the participants are at all engaged with one another's thinking. This kind of talk is not

uncommon in primary classrooms and illustrates a need for the guided development of children's skills in communicating and thinking together. In Transcript 2 the children's discussion has many of the features of Exploratory Talk. They ask each other for information and opinions, seek reasons and provide them, share their thoughts, and evaluate proposals that are made. Challenges are constructive and all the group are involved in working towards a joint decision. The children's use of the structures they have recently been taught may seem a little formulaic (for example, Alana's repeated 'Why do you think that?'); but they are using the strategies appropriately and purposefully. Opinions are treated with respect. Each speaker has the opportunity to develop their ideas as they speak. Indeed, it appears that the talk precipitates thinking. The two researchers who carried out the 'blind assessment' exercise correctly identified the transcribed discussions correctly as 'pre-' and 'post-intervention' for five out of the six classes.

Quantitative analysis of talk

Table I shows the relative the incidence of indicator words in target children's talk while doing *Kate's Choice*, before and after their involvement in the Thinking Together lessons.

<insert Table I here>

It can be seen that there is a greatly increased incidence of the words indicative of Exploratory Talk in the children's discussions after the intervention. The children's greater use of such words reflects their increased attempts to share each other's thoughts before deciding on a course of action and moving on through the programme. In earlier research, we had found that the more children explain and

justify their views, the longer their utterances tend to be. The association between increased utterance length and more explanation and justification was confirmed in the present study by the qualitative analysis of a sample of transcripts. We therefore also counted the number of utterances in all the talk data which exceeded 100 characters when transcribed (excluding text read from the computer screen). The results of this analysis, provided in Table II, are indicative of the more elaborated contributions to the discussions made by children post-intervention.

<insert Table II here>

Evaluating knowledge and understanding of the science curriculum

The results of the assessment based on science SATs questions are shown in Table III. The numbers of children are less than the total number who participated, because to ensure validity of statistical analysis it was necessary to exclude those children for whom no pre/post match was possible due to departures during the project and absences at the times of testing. It was also unfortunate that because of staffing problems in one school at the time of the post-intervention assessment, one class had to be excluded. Nevertheless, the results indicate that the scores of target classes increased significantly more than those of the control classes. This supports the view that the intervention had a positive effect on the target children's study of the relevant parts of the science curriculum.

<insert Table III here>

The concept mapping exercise was used as a test of the extent to which children in the target classes became, over the period of the intervention, relatively more able to

perceive relationships between different scientific concepts - or at least the terms representing them. The results of the exercise are summarised in Table IV, which indicates the impact of the intervention in the predicted direction. Because the intervention did not involve any special teaching about the relationships between concepts or terms, the target children's superior post-intervention performance in producing concept maps seems most appropriately attributed to the improved quality of their collective reasoning about the relevant subject matter.

<insert Table IV here>

Overall, the results show that the children in the target classes gained significantly better scores in science than those in control classes, thus providing evidence for the effectiveness of the intervention in improving children's study of the science curriculum.

Changes in the quality of collective and individual reasoning

The mean scores of children carrying out the Raven's test are shown in Table V below. Pre- and post- intervention comparisons of groups of children in the target and control classes working on Raven's test were made using an analysis of covariance with pre-intervention results as covariate, post-intervention results as dependent variable and condition as fixed. It can be seen that the target pupils performed significantly better than the control pupils after the intervention, taking into account the pre-intervention performance levels of both groups.

<insert Table V here>

In the school environment, some changes in the pre/post membership of groups were

inevitable. Such changes could be assumed to reduce the covariance between the pre- and post-test scores, which would mostly decrease the accuracy of the corrected post-score measures (in the analysis of covariance) by introducing ‘noise’. This would normally make it harder to detect a significant difference due to the treatment condition. It is possible the group composition changes may not have been random and could have introduced bias into the analysis of covariance: but this cannot be detected or corrected. Taking these two considerations into account, the results for the groups are strongly indicative of the intended effect of the intervention.

Pre/post comparisons of individual children in the target and control classes doing the Raven’s test were also made using an analysis of covariance with pre-intervention results as covariate, post-test results as dependent variable and condition as fixed. The mean scores are shown in Table VI below. The target children performed significantly better than the control children after the intervention, taking into account pre-intervention performance levels.

<insert Table VI here>

As mentioned earlier, our previous studies had shown statistically significant increases in individual scores on this test, but not in corresponding group scores. These new results provide statistical evidence of changes in both group and individual reasoning. They thus provide support both for the general sociocultural hypothesis that the intermental activity of using language as a tool for reasoning collectively can influence the development of individual thinking and learning. They also support the more specific hypothesis that a programme of activities for encouraging the use of Exploratory Talk had specific, predicted positive effects on the quality of children’s

reasoning. Moreover, these effects are demonstrated by a test which was not similar in content or format to any activities in the Thinking Together programme, meaning that the target children were no more familiar with the test than the control children. This therefore also supports the view that the discursive and reasoning skills gained by the target children were not highly task-specific or context-bound, but represented a transferable competence. By involvement in a class undertaking the teaching and learning of talk skills, the children had gained the kinds of generalisable communication and thinking skills which are commonly advocated as a desirable outcome of educational experience.

In order to evaluate longer term effects sample 'spot check' video recordings of activities with children from two target schools have been made, almost one year after the end of the intervention. These showed that individuals and groups were still able to recall the ground rules and use Exploratory Talk appropriately when carrying out problem-solving activities. One of our intentions has been to follow Adey and Shayer's (1994) example by post-testing for the impact of the intervention in future years, by tracing target classes on measures of academic attainment. However, as mentioned earlier, there are high rates of mobility of participating children and teachers in Milton Keynes and this makes it difficult to track any one cohort of children. There is also the problem of the lack of continuation of the Thinking Together approach as participating teachers leave a school or as children move to other schools, with the likely result that the discourse habits which the approach has encouraged in children will not be reinforced. Nevertheless, our continued involvement with Milton Keynes schools should make it possible to assess continuing effects on practice and achievement.

Discussion and conclusions

The research reported here has demonstrated that an experimental teaching programme enabled children in primary schools to work together more effectively, improve their language and reasoning skills and reach higher levels of attainment in their study of science. We have replicated and strengthened our earlier findings which supported the sociocultural claim that language-based, social interaction (intermental activity) is a developmental influence on individual thinking (intramental activity). But the results provide support of other kinds for the sociocultural perspective on education, and hence for the validity of sociocultural theory as the basis for an applied psychology of education. By showing that children's increased use of certain ways of using language leads to better learning and conceptual understanding in science, we have provided empirical support for the conception of science education as induction into a community of discourse or practice (Lemke, *op.cit.*; Leach and Scott, *op.cit.*).

The intervention programme was carefully designed to include both teacher-led activities and activities in which children worked together in groups; and as the teachers in target schools were trained to provide guidance of a particular kind, and the group discussion in target schools was expressly based on 'ground rules' which were not made available to control classes, our study does not enable the effects of these two factors to be distinguished. But our theoretical position supports the view that separating these two types of interaction is incompatible with the provision of an effective educational experience; it would not be possible to design an intervention compatible with our sociocultural hypotheses without combining these two elements. Moreover, the complementary effects of both types of interaction for primary science education have been well demonstrated by other very relevant research (particularly

that of Howe et al., 2000, as discussed in the earlier part of this paper). Our findings therefore add to the evidence that the development of scientific understanding is best assisted by a careful combination of peer group interaction and expert guidance, and provide an example of how that combination can be successfully achieved.

These findings have relevance for current debates about the role of educational research for providing a base of evidence for informing educational practice. They demonstrate that the pursuit of applied, practical educational research need not entail the loss of a theoretical dimension in favour of mere evaluative description.

Theoretically-informed ideas about teaching and learning can be transformed into empirical investigations, which can then provide new insights into ‘what works’ – and thus provide an evidential base for practice of the most robust kind. Surely practice is best informed if it has both theoretical and empirical support?

Evidence of the positive effects of our intervention on children’s talk and attainment has been provided by several, complementary strands of our analysis. However, there are also other observed effects, more subtle and harder to measure precisely, on ways teachers work with their classes, on groups undertaking joint activities, and on individual children. The teachers in the project schools have created talk-focused classrooms in which the children, as Barnes and Todd (*op. cit.*) put it, are ‘each other’s resource’. But this does not mean teachers should abdicate responsibility for guiding the construction of knowledge; their role in enabling children to gain a better understanding of interpersonal communication and curriculum content is crucial. Rather, it creates conditions in which the educational purposes of teacher-led and

group-based activities become clearer to all involved. Teachers report that their relationship with their class benefits, as the class gains an ethos based on shared purposes for activity and, especially, for collaboration. The class atmosphere becomes more open, interested and engaging.

Providing children with ‘rules’ for talk may seem constraining. But for children who agree some ground rules and then try to implement them, they can represent a kind of freedom. The usual social conditions for talk – for example, the dominance of the participants who talk most and most forcefully - are suspended. The social status of individuals can be neutralised by the ground rules, creating an intellectual environment which is more equitable – though of course it is also one in which everyone’s ideas are open to critical examination. More confident children gain the opportunity to hear a wider range of views. Quieter children find that their contribution is sincerely requested and valued. One of the simplest but most profound benefits for children is the idea that challenging each other is not just accepted but encouraged. In many situations, having your views challenged can be threatening. In the target groups, however, a challenge leads to an exchange of reasoning and a better understanding of another’s position. The insight that this is possible may be generalised to other situations. Teachers have reported that participating children find it easier to resolve conflict in situations outside the classroom. Our findings indicate that if teachers provide children with an explicit, practical introduction to the use of language for collective reasoning, then children learn better ways of thinking collectively and better ways of thinking alone. The implications for educational practice are clear.

In order that the results of the project can inform educational practice, a set of lesson

plans called *Thinking Together in Maths and Science* has been created and will be published as a companion volume to our earlier publication (Dawes et al., *op.cit.*). This will incorporate guidelines for the effective use of existing educational software as a basis for group activity. Draft multimedia material for the professional development of teachers on the implementation of the Thinking Together approach has also been produced. It sets out the structure of the Thinking Together programme and the teaching strategies involved in implementing it, using video sequences of classroom interaction from the project data to illustrate ways of modelling and scaffolding children's questioning and reasoning. This has already been used to conduct professional development sessions for schools and workshops at conferences and plans are now under way to use it as the basis for an Open University on-line in-service course for teachers. Related projects, building on the methods and results described here, have begun in Milton Keynes schools for Key Stages 1 (ages 6-7) and 3 (ages 12-13).

In the UK, findings from our earlier research have already informed the official guidance and training materials provided for teachers in relation to English and the national strategies in literacy and the foundation subjects (e.g. DfES, 2002; QCA, 2003) and have been presented to a House of Commons Select Committee on Education and Skills. However, across the curriculum subjects there is considerable variation in terms of official policy, teacher guidance and classroom practice in the extent to which it is recognized that the quality of teaching and learning is dependent on the quality of classroom dialogues. It is our hope that the findings reported here will encourage greater recognition of this in the teaching of science.

References

- Adey, P. and Shayer, M. (1993) An exploration of long-term far-transfer effects following an extended intervention programme in the high school science curriculum, *Cognition and Instruction*. 11 (1), pp. 1-29.
- Alexander, R. (2003) Talk in teaching and learning: international perspectives, in *New Perspectives on Spoken English in the Classroom: discussion papers*. London, Qualifications and Curriculum Authority.
- Barnes, D. and Todd, F (1977) *Communication and Learning in Small Groups* (London, Routledge & Kegan Paul).
- Daniels, H. (2001) *Vygotsky and Pedagogy* (London: Routledge/Falmer).
- DfES (2002) *Training Materials for the Foundation Subjects: Key Stage 3 National Strategy* (London, Department for Education and Skills).
- Galton, M. & Williamson, J. (1992) *Group work in the Primary Classroom* (London, Routledge).
- Harrison, C., Comber, F., Fisher, T., Haw, K., Lewin, C., Lunzer, E. McFarlane, A., Mavers, D., Scrimshaw, P., Somekh, B. & Watling, R. (2002) The Impact of Information and Communication Technologies on Pupil Learning and Attainment (The ImpaCT II study). *ICT in Schools Research and Evaluation Series No. 7*. Coventry: Becta /DfES. [www.becta.org.uk/research/reports/docs/ImpaCT2_strand1_report.pdf]
- Howe, C., Tolmie, A., Duchak-Tanner, V. and Rattray, C. (2002) Hypothesis testing

in science: group consensus and the acquisition of conceptual and procedural knowledge, *Learning and Instruction*, 10, 4, pp. 361-91.

Leach, J. & Scott, P. (1995) The demands of learning science concepts: issues of theory and practice, *School Science Review*, 76, pp. 47-51.

Lemke, J. (1990) *Talking Science: language, learning and values*, (Norwood, N.J: Ablex).

Mercer, N. (1995) *The Guided Construction of Knowledge: talk amongst teachers and learners* (Clevedon, Multilingual Matters).

Mercer, N., Wegerif, R. & Dawes, L. (1999) Children's talk and the development of reasoning in the classroom, *British Educational Research Journal*, 25, 1, pp. 95-111.

Ollerenshaw, C. & Ritchie, R. (1998) *Primary Science: making it work*, (London, David Fulton Publishers).

QCA (2003) *New Perspectives on Spoken English in the Classroom: discussion papers*. London, Qualifications and Curriculum Authority.

Raven, J., Court, J. and Raven, J. C. (1995) *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Oxford, Oxford Psychologists Press).

Rogoff, B. (1990) *Apprenticeship in Thinking* (New York, Oxford University Press).

Rojas-Drummond, S., Mercer, N. & Dabrowski, E. (2001) Collaboration, scaffolding and the promotion of problem solving strategies in Mexican pre-schoolers, *European*

Journal of Psychology of Education. XVI (2), pp. 179-96.

Rojas-Drummond, S. & Mercer, N. (in press) Scaffolding the development of effective collaboration and learning, *International Journal of Educational Research*. (Forthcoming special issue on Group Work, edited by P. Blatchford & P. Kutnick.)

Russell, T. and Watt, D. (1990) *SPACE (The Science Process and Concept Exploration Project) Research Report: Growth* (Liverpool, Liverpool University Press).

Vygotsky, L. S. (1978) *Mind in Society: the development of higher psychological processes* (Cambridge, Mass.: Harvard University Press).

Wegerif, N., Mercer, N. and Dawes, L. (1999) From social interaction to individual reasoning; an empirical investigation of a possible socio-cultural model of cognitive development, *Learning and Instruction*, 9, pp. 493-516.

Wegerif, R. and Mercer, N. (1997) Using computer-based text analysis to integrate quantitative and qualitative methods in the investigation of collaborative learning, *Language and Education*, 11, 4, pp. 271-87.

Wegerif, R. and P. Scrimshaw (1997) (Eds.) *Computers and Talk in the Primary Classroom* (Clevedon, Multilingual Matters).

Wegerif, R. and Mercer, N. (in press) The 'talking bug': using a computer-based tool to guide discourse in collaborative group. *Hermes*.

Wells, G. (1999) *Dialogic Inquiry: towards a sociocultural practice and theory of education* (Cambridge, Cambridge University Press).

Wertsch, J. (1991) *Voices of the Mind: a socio-cultural approach to mediated action* (Cambridge, Mass.: Harvard University Press).

Wood, D., Bruner, J. and Ross, G. (1976) The role of tutoring in problem-solving. *Journal of Child Psychology and Child Psychiatry*, 17, 89–100.

Acknowledgements

The researchers gratefully acknowledge the support of the Nuffield Foundation (Project EDU/00169/G) and the participation of teachers and children in Milton Keynes schools. We also register our appreciation of the expert advice and assistance of Dr Martin Le Voi of the Open University in carrying out and verifying the statistical analysis.

Table I: Relative incidence of indicator words: Kate's Choice

Indicator word	Pre-intervention talk in focal groups in target classes	Post-intervention talk in focal groups in target classes
because	13	50
I think	35	120
would	18	39
could	1	6
Totals	65	215

Table II: Relative incidence of long utterances during *Kate's Choice*

	Pre-intervention talk; target classes	Post-intervention talk; target classes
Utterances 100 + characters	1	46

Table III: Results for the SATs science questions

	Numbers	Pre-intervention: mean scores	Post-intervention: mean scores
Target classes	119	3.97	5.70
SD		2.323	2.424
Control classes	129	4.22	5.04
SD		1.997	2.206
Effect size	.29		

$F(1, 245) = 10.305$; two-tailed $p = 0.002$

Table IV: Results of the Science Concept Mapping Exercise

	Numbers	Pre-intervention: mean scores	Post-intervention: mean scores
Target classes	115	1.52	6.02
SD		1.455	3.364
Control classes	129	1.08	3.97
SD		1.473	2.756
Effect size	.74		

$F(1, 241) = 17.471$; two-tailed $p = 0.000$

Table V: Group performances of children in Target and Control classes on the Raven's test

	Pre-intervention: mean scores	Post-intervention: mean scores
Target classes	20.08	23.62
SD	4.052	2.413
Control classes	19.90	22.36
SD	3.423	2.290
Effect size	.55	

$F(1,76) = 6.281$; two-tailed $p = 0.014$

Table VI: Individual performances of children in Target and Control classes on the Raven's test

	Pre-intervention: mean scores	Post-intervention: mean scores
Target classes	16.26	18.90
SD	3.981	3.542
Control classes	15.60	17.88
SD	4.357	3.702
Effect size	.27	

F (1,224)=6.065; two-tailed p = 0.015