

**Using Exploratory Talk to Enhance Problem-solving and Reasoning Skills in
Grade-7 Science Classrooms**

Paul Webb

Department of Science, Mathematics and Technology Education
Nelson Mandela Metropolitan University
Port Elizabeth 6001, South Africa
Phone: +27-41-836516991
Fax: +27-41-5821000
Email: paul.webb@nmmu.ac.za

David F Treagust

Science and Mathematics Education Centre
Curtin University of Technology
Perth WA 6845, Australia
Phone: 61 8 9266 7924
Fax: 61 8 9266 2503
Email: D.Treagust@smec.curtin.edu.au

**Using Exploratory Talk to Enhance Problem-solving and Reasoning Skills in
Grade-7 Science Classrooms**

ABSTRACT

This study investigates whether the generation of exploratory talk in grade seven, second-language science classrooms in the Eastern Cape Province, South Africa, has a positive effect on learners in terms of their problem-solving and reasoning skills and whether socio-cultural *milieus* (urban, peri-urban and rural settings of schools) have an effect on any cognitive gains that may be made by learners. The findings were that there is a clear and statistically significant improvement in the mean test scores on problem-solving and reasoning of pupils who participated in the classroom discussion initiative (exploratory talk) over those of the comparison groups. This result was consistent in both the first- and second-studies that were undertaken. There was also a significant difference between the gains in problem-solving and reasoning test scores of the three socio-geographical settings in which the participating schools are found.

Using Exploratory Talk to Enhance Problem-solving and Reasoning in Grade-7 Science Classrooms

INTRODUCTION

In recent years, psychological researchers such as Bruner (1990), Rogoff (1990), Rogoff, Gauvain and Ellis (1991) and Wertsch (1991) have become increasingly concerned with understanding how children's thinking is shaped by social experience amongst peers and by adult guidance. Building on the work of Vygotsky (1962) the researchers have elaborated a sociocultural theory of intellectual development in which the social experience of language use is seen as a major shaper of cognition (Mercer, Wegerif & Dawes, 1999). Similarly, constructivist research (particularly notions of social constructivism) has sensitised science educators to the importance of classroom discussion (Sprod, 1995), and Solomon (1994) has gone as far as to champion a change in metaphor in science education to that of a child on the edge of a circle of initiates trying both to make sense of the conversation and be accepted into it.

However, observational studies of classroom life rarely reveal any evidence of systematic induction of children into ways of using language for seeking, sharing and constructing knowledge (Mercer, Wegerif & Dawes, 1999). In fact, researchers have found that children commonly lack any clear, shared understandings of the purposes of many of the activities in which they are engaged and so are often confused, unfocused and unproductive in their use of language (Edwards & Mercer, 1987; Sheeran & Barnes, 1991; Barnes & Todd, 1995). Some researchers have therefore concluded that the educational and developmental potential of classroom conversation (particularly amongst pupils) is being squandered (Galton & Williamson, 1992; Christie & Martin 1997). This

appears to be particularly true in South Africa as there is little evidence of meaningful discussion in the classrooms of schools which were previously disadvantaged under the system of Apartheid, and where both teachers and learners officially operate in their second language (Taylor & Vinjevold, 1999). However, children are no longer segregated in terms of race and the Revised National Curriculum Statement, which is underpinned by notions of constructivism, requires that teachers promote environments where both learners and teachers can interact, discuss and exchange ideas (Department of Education, 2002).

Wegerif, Mercer and Dawes (1999) have shown that children's test scores on the Raven's Standard Progressive Matrices test (a standardised psychological test of non-verbal reasoning) improved significantly after they had been taught to use 'exploratory talk' (after Barnes & Todd, 1978) during group activities aimed at solving the reasoning problems posed in the Raven's test. These data have led them to argue that as individual reasoning has part of its origin in dialogue with others, the experience of social reasoning played a significant role in the improved scores on measures of individual reasoning that were recorded (Wegerif, Mercer & Dawes, 1999). They note, however, that although they have some qualitative evidence that children are able to apply the communicative reasoning skills that they develop to the study of the science curriculum, this remains a crucial area for attention in future research (Mercer, Wegerif & Dawes, 1999).

This study was guided by two research questions: 1. Can teachers be taught to generate curriculum based exploratory talk in their classrooms such that the ensuing discussion has an effect on learner's cognitive development in terms of their problem-solving and reasoning skills?, and 2. Are the different socio-cultural *milieus* in which

previously disadvantaged South African schools are set reflected in the type of discussion that is generated and in the degree of cognitive development that takes place?

BACKGROUND

Discussion and reasoning

Whereas traditional psychology has described reasoning in terms of logical rules, a number of arguments in contemporary cognitive psychology point to the need to see reason as a form of social practice (Resnick, Salmon, Zeitz, Wathen & Holowchak, 1993). Wegerif, Mercer and Dawes (1999) argue that as Vygotsky stresses that “all that is internal in the higher mental functions was at one time external” (Vygotsky 1981, p. 36), his influential model of individual development presupposes and stems from prior socialisation processes. Leont’ev (1981) refers to the development that Vygotsky called ‘internalisation’ as a process of the personal appropriation of shared cultural capital that results from a period of guided participation or cognitive apprenticeship. A number of authors have followed this train of thought and support a dialogical account of reasoning which implies that reasoning is embedded in a social practice (Newman, Griffin & Cole, 1989; Rogoff, 1990; Rojas, Drummond, Hernandez, Velez & Villagran, 1998).

Classroom discourse patterns

Classroom discourse patterns have been studied fairly extensively since Flanders’ (1970) early work (see for example, Mortimer & Scott, 2000) and the dominance of Sinclair and Coulthard’s (1975) Initiation-Response-Feedback (IRF) questioning cycle has been

confirmed by Lemke (1990). This technique (IRF) has been criticised as being rooted in the need for teacher control of the discussion, rather than an educational rationale, which leads to shallow thinking and a ‘guess what the teacher is thinking’ approach to learning (Edwards & Mercer, 1987; Dillon, 1994). In turn, this promotes an epistemological understanding in learners that scientific knowledge is fixed, revealed and uncontested (Lemke, 1990). By contrast, ‘true dialogue’ and ‘cross-discussion’ are two discourse patterns that Lemke (1990) identifies as best supporting constructivist learning. However, these are the discourse patterns least used in classrooms around the world (Lemke, 1990).

Mercer (1996a, 1996b) used observational research in British primary schools to typify three kinds of classroom talk, which he also described as different modes of social thinking. These modes of talk and, according to Mercer (1996a), of thinking are *disputational* talk, *cumulative* talk and *exploratory* talk. Disputational talk is characterised by disagreement and individualised decision-making, with few attempts to pool resources or to offer constructive criticism or suggestions and notably consists of short exchanges consisting of assertions and counter-assertions. Cumulative talk occurs when speakers build positively, but uncritically, on what others say and, as such, cumulative discourse is characterised by repetitions, confirmations and elaborations. Exploratory talk is when partners engage critically but constructively with each other’s ideas, statements and when suggestions are offered for joint consideration and challenges and counter-challenges are justified and alternative hypotheses are offered. Compared to the other two discourse types, in exploratory talk knowledge is made more publicly accountable and reasoning is more visible in the talk, while progress emerges from the eventual joint agreement reached (Mercer, 1996a).

It is talk of an exploratory nature, where relevant information is shared, the group seeks agreement and takes responsibility for decisions, reasons and challenges are expected and accepted, alternatives are discussed and all members of the group are encouraged to speak, that Wegerif et al. (1999) believe is the socio-linguistic process that improves group and individual reasoning in children.

Promoting discussion in science classrooms

There are a number of models for promoting effective discussion (Dillon, 1994). For example the 'Philosophy for Children' programme (Lipman, Sharp & Oscanyan, 1980) used purpose-written stories with open-ended puzzles to trigger discussion. The discussion is then based on questions that the children ask and the role of the teacher is to require good thinking about the questions (usually by modelling and cuing). Gardner (1995) asserts that these types of discussions are neither learner-centred nor teacher-centred: they are 'truth-centred'.

In this study, the hypothesis was made that trained teachers could initiate and sustain discussion of an exploratory nature in their science classrooms by using specially developed 'hands-on' practical work activities, shared reading of text (the 'Big Book') and question and challenge-based prompt-posters, and that the ensuing discussion would have an effect on learners cognitive development.

It was also hypothesised that the different socio-cultural *milieus* in which schools are set, *viz.* urban, peri-urban and rural settings, may be indicative of learners' 'cultural capital' (after Vygotsky) and that this might be reflected in the type of discussion

generated. In turn, it was hypothesised that the type of discussion generated would affect the degree of cognitive development that takes place.

Worksheets, Big Books and Prompt Posters

The teachers who participated in this study were provided with the apparatus and written materials (worksheets) necessary for small groups of their pupils to complete certain science curriculum-based (magnetism) practical tasks and given big-books and prompt posters on the same curricular topics. The teachers also attended workshops that emphasised the nature of exploratory talk and allowed them to practise strategies to initiate and maintain talk of this nature while using the materials provided.

The worksheets and apparatus that were provided enabled activities and discussion on polarity, using magnets to find direction, magnetic fields and looking after magnets. The big-book on magnetism was designed for shared reading by groups and firstly involved the pupils in pre-reading activities in which they discussed, amongst others, what they thought 'a navigator' is and how sailors find their way across the oceans. Thereafter, the pupils were encouraged to discuss the story at regular intervals and to respond as groups and individuals to specific issues such as why they thought that magnets lose their magnetism if dropped or heated and keepers help magnets maintain their magnetism while stored.

The prompt poster is a large, folded card that is draped over a holdall-lectern. The side facing the class presents an illustration of an event (such as various ways of finding direction by means of a magnet) to the class on a large (double A3) surface. The reverse side facing the teacher has a number of prompts (suggested questions and challenges for

the teacher to use) to help the pupils focus on specific aspects of the visual and to promote dialogue.

Raven's Standard Progressive Matrices

Cross cultural studies suggest that the maturation of intellectual capacity is partly the result of environmental influences and cultural opportunities, at least to the extent that in the absence of stimulation the development of logical thinking tends to remain latent, or to develop somewhat later in life (Raven, Court & Raven, 1995). Raven's Progressive Matrices (RPM) consist of graphical puzzles and are widely used in education and psychology as a test of the ability to reason and solve problems involving new information and, as such, are seen as an indicator of the capacity for systematic reasoning and logical thinking (Carpenter, Just & Shell, 1990).

Wegerif et al. (1999) argue that the Raven's tests are particularly appropriate for exploring the link between language practices and the non-culturally based tradition of research in cognitive development because they correlate well with other tests of reasoning and with measures of academic achievement (Raven, et al, 1995; Richardson, 1991). Carpenter et al. (1990) also suggest that this centrality of the Raven test indicates that it is not only a good measure of intelligence, but that the processing required in the Raven test should account for a good deal of the reasoning in the other tests.

Richardson (1991) notes that standardised, non-verbal reasoning tests like Raven's have commonly been taken to be paradigmatic measures of individual reasoning ability, independent of social or cultural factors. However, he argues that RPM do not measure abstract mental processes but rather the ability to read a particular kind of representation

and has demonstrated that if the same logical problems as those found in Raven's tests are presented using pictures of cars or teddy bears instead of abstract shapes, children respond very differently with a different distribution of test scores. Despite the cultural dependency implied by Richardson (1991), the Raven's tests are language-free and the use of abstract shapes as required is a valued kind of cognitive ability in terms of many of the processes used in science (Wegerif et al., 1999).

Dix (1998) used the Raven's Standard Progressive Matrices Test (RSPM) to obtain a measure of mathematical reasoning ability as it provides a measure of educative ability or fluid intelligence which is relatively independent of specific learning acquired in a particular cultural or educational context and that it can be used to provide a common base-line measure between students and classes (de Lemos, 1989).

In the context of this study, the Raven's Standard Progressive Matrices (RSPM) test was chosen as it is a well-established, reliable test of the ability to reason and solve problems using new information, it can be used across a range of ages and correlates highly with measures of academic achievement (Carpenter et al., 1990), and it has been used in similar fields of study (Dix, 1998).

The Raven's Standard Progressive Matrices are used to assess current capacity to comprehend graphical puzzles, and the ability to perceive the relationship between them and is, according to Raven et al. (1995), suitable for children from the age of six years to adulthood regardless of language, education, nationality or physical condition. Also Raven's Progressive Matrices are used predominantly as a culture-free test. All of these factors were deemed to be particularly important in the context of attempting to test

children from a Xhosa culture and largely deprived backgrounds, and who are instructed at school in a second language.

The Raven's tests appear particularly appropriate for exploring links between language practices and the non-culturally biased tradition of research in cognitive development as they correlate well with similar tests of reasoning and with measures of academic achievement (Raven et al., 1995; Richardson, 1991). Further, Raven's tests correlate with measures of information processing (Bereiter & Scardamalia, 1979). The Raven's tests are, however, not tests of general intelligence and should preferably be used in conjunction with a test of vocabulary, i.e. the Crichton Vocabulary Scale (Raven et al., 1995), but this additional vocabulary scale is not standardised for use in South Africa and, as such, was not used in this study.

The Raven's Standard Progressive Matrices (RSPM) is divided into five sets of twelve problems (sets A, B, C, D and E). Each set starts with a problem which is, as far as possible, self-evident, and develops a theme in the course of which the problem builds on the argument of what has gone before and thus becomes progressively more difficult. The cyclical format provides an opportunity to assess the consistency of a person's intellectual activity across five successive lines of thinking (Raven et al., 1995).

Raven et al. (1995) note that a person's maximum capacity for clear thinking has been found to vary with health and to improve with practice, and varies less with a person's speed of accurate intellectual work. Therefore, for anthropological, genetic and clinical studies, an untimed 'capacity' test is deemed to be more useful than a test in which a person has to work against the clock. As such, timing of the testing process was

not a factor in this study and all of the children were allowed to complete the exercise (+/- 45 minutes). This period seemed to be quite suitable to assess the children's capacity for coherent perception and orderly judgement without causing exhaustion.

Where reading levels are adequate (grade 3 and above), the Raven's Standard Progressive Matrices tests can be administered in book form with answer sheets. However, the booklets, though re-usable, are extremely expensive in South African terms and consequently whole-class testing in this form has serious financial constraints. This difficulty was overcome by using overhead projector (OHP) transparencies displayed sequentially. The pupils were required to indicate completion of each test item by putting down their pens, giving the tester an indication of when to move on to the next item.

As our interest in this study is to investigate the relationship between cognitive ability and the way children talk together, the Raven's test was judged to be an appropriate, valid and cost effective means of researching this issue in the South African context.

The South African Context

There is broad consensus that teaching and learning in the majority of South African schools leaves much to be desired (Taylor & Vinjevold, 1999). The problems are generally described in terms of teacher-centredness, pupil passivity and rote learning. A number of researchers and organisational reports (ANC, 1994; Chisholm, 1993; Enslin, 1990; Hartshorne, 1992; Hofmeyr, 1993; NEPI, 1992) suggest that the ideology of Fundamental Pedagogics is to blame and has had wide ranging detrimental effects on teachers' thinking and practice. (Fundamental Pedagogics is an indigenous South African

product drawing on Dutch phenomenological philosophy that claimed to have developed a science of education.)

Fundamental Pedagogics is based on premises that can be interpreted as being authoritarian (e.g., the teacher, as knowing adult, leads the child to maturity). However, it was more the way in which it was taught, *viz.* through a series of propositions that brooked no analysis or critique, that produced the detrimental effects on teachers' thinking and practice. Enslin (1990) argues that Fundamental Pedagogics heads off the possibility of critical reflection by making reflection illegitimate by justifying authoritarian practices. As such, she describes Fundamental Pedagogics as an ontology that produces (in terms of promoting Apartheid aims) useful and docile teachers.

Macdonald (1990, 1991) found that black children spent most of their time in class listening to their teachers and that the dominant pattern of classroom interaction was oral input by the teacher with the children occasionally chanting in response. Teachers did ask questions, but these were aimed at data recall or checking whether the children were listening to the lesson rather than eliciting more challenging responses. Classroom tasks in general were aimed at the gaining of information rather than higher cognitive tasks.

Research that formed part of Nelson Mandela's *Presidential Education Initiative* (PEI) confirms the findings of Macdonald (Taylor & Vinjevold, 1999). PEI researchers found that the majority of questions posed by teachers usually involved simple data recall, or were simply used to test whether the pupils were listening, and where there were instances of more difficult questions being asked by teachers they were answered in a sing-song chorus, suggesting that the answers were learnt by heart. These findings were confirmed during pre-research visits to the schools that participated in this study and, as

such, the assumption was made that prior to this intervention little to no discussion, which could be described as ‘exploratory’, took place in the experimental or comparison classrooms.

Despite ambitious public commitment by the South African government to the provision of high quality and progressive learning materials, and the recognition by the international literature that some of the most important predictors or precursors of cognitive development is the access of pupils to learning materials such as books and stationery (Crouch & Mabogoane, 1997), many (if not most) South African schools do not receive the materials they need (Taylor & Vinjevold, 1999).

The dismantling of Apartheid legislation saw rapid diversification of previously linguistically homogenous schools in South Africa. This is especially true of urban and peri-urban schools in the townships and squatter camps near big towns and cities. As such, Brown (1998) concludes that because of rapidly changing demographics a significant proportion of South Africa’s learners will face a situation where their home language is not on offer at the schools that they attend. Also, since the 1950s black African parents have opposed mother tongue instruction as it was seen as a strategy by the government to prevent African upward mobility and thereby ensure a perpetual reservoir of cheap labour (NEPI 1992). This has resulted in a situation where, for the vast majority of South African schools, all teaching and learning takes place in a second language (English), despite the fact that policy makers unequivocally support mother tongue instruction, and that many researchers report on the positive effects of mother

tongue instruction, at least in the early years of schooling (Rodseth, 1995; Taylor & Vinjevold, 1999).

In the light of the above, the teachers who participated in this project were provided with materials and apparatus and engaged in workshops to practice strategies which included code switching between mother tongue and English in an attempt to maximize the possible effects of discussion on the cognitive development of their pupils. Observation of the teachers in these workshops revealed that they were able to engage in code switching (an activity that is regularly observed in South African classrooms where instruction is in the learners' second language) and the other strategies aimed at promoting effective discussion.

DESIGN

This research was carried out using two complementary studies. A first study over a six-month period was carried out in 12 schools in three *milieus* or settings in 2001. The purpose of the first study was to elucidate, *via* quantitative testing and qualitative classroom observations and interviews, whether the generation of exploratory talk in grade 7 second-language science classrooms in the Eastern Cape Province, South Africa, had positive effects on learners in terms of cognitive development and whether socio-cultural *milieus* (urban, peri-urban and rural settings of schools) had an effect on the discussion generated and the cognitive gains made by learners.

The second study in 2002 also operated over a six-month period in 12 new schools in three *milieus* or settings in schools that were matched by using the same set of criteria as used in the first study. However, as the first study indicated no significant differences between the methods employed to initiate discussion, in the second study all three methods were used in combination in all treatment schools. Both quantitative and qualitative techniques were used to generate and analyse the data generated.

Materials

A series of matching ‘triggers’ (activities and strategies aimed at promoting discussion and ‘triggering’ discussion between learners) focusing on magnetism were developed for grade 7 classrooms. Magnetism was chosen as it is topic in the grade 7 curriculum and provides a setting that allows both fairly simple practical activities and ideas that can be discussed at both superficial and highly conceptual levels. The triggers were practical activities (doing science using a set of worksheets and apparatus provided to teachers), conversational readings (engaging in stories that illuminate science in everyday activities), and ‘prompt posters’ (discussion about instances or situations). Each set of trigger material covered four aspects of magnetism, *viz.* ‘Magnetic and non-magnetic materials’, ‘Magnetic fields and forces’, ‘Finding direction using magnets’ and ‘Caring for your magnets’.

Participants

Twelve grade-7 science teachers, selected to participate in the first study, were matched by their success as students in the University of Port Elizabeth BEd (Science and

Mathematics) programme that operates in Port Elizabeth, KingWilliam'sTown and Queenstown (i.e., urban, peri-urban and rural geographic and socio-cultural *milieus*) and were allocated (randomly in each location) to a trigger or comparison group (one teacher and one class) in separate schools. As such, each trigger was applied in a classroom in each location and there was a comparison group in each location. Comparison group teachers were identified at the beginning of the study and, subsequently played no further role in workshops and other classroom discussion activities. They were only visited at their schools in order to observe their science classroom activities, i.e. when teaching magnetism.

The schools were broadly matched in terms of being chosen as institutions that were neither dysfunctional nor excellent, were from disadvantaged communities and had grade 7 class sizes between 30 and 50 learners. The schools were chosen so that they could be grouped equally (i.e., four per group) in urban, peri-urban or rural settings. Of the four teachers identified in each *milieu* (setting), one school (i.e. the teacher plus his/her class of pupils) was randomly chosen at the start of the study to act as a comparison group in order to allow comparisons with the experimental groups while each of the other three teachers were allocated a set of trigger material. As noted earlier, the comparison group teachers played no further part in the 'teacher development' activities with respect to the 'treatment' and remained unaware of the classroom discussion activities that were being promoted in the other classrooms. The second study over a six-month period in 12 new schools had a similar arrangement except that the treatment teachers used a combination of the three types of triggers.

Training

The teachers comprising the experimental group were introduced to notions of classroom discussion and trained via workshops in the use of the specific trigger materials to which they had been randomly allocated. Firstly, all participating teachers (other than the comparison group) attended a one-day workshop on '*the place of talk in science classrooms*' where they were introduced to possible strategies and criteria for establishing classroom discussion. These strategies included ensuring their learners were aware of the structure of good whole-class discussion, e.g. that interactions could be teacher to learner, learner to teacher, or learner to learner, and that these interactions could be observations, explanations or questions. Rules of politeness applied, and the 'game strategy' was to enable as many types of interaction to take place during a lesson. The danger of slipping into an Initiation-Response-Feedback (IRF) mode was also emphasised. A second one-day workshop (the following week) was held where the experimental group of teachers were provided with the appropriate trigger material materials, as well as information and assistance in terms of using the materials and sequencing the activities. The comparison teachers had the teaching materials delivered to their schools (the apparatus, readings and prompt posters), but were not part of the workshops aimed at promoting classroom discussion.

Measures

Before the teachers applied the triggers in their classroom, measures of their learners' reasoning skills were taken using the Raven's *Standard Progressive Matrices* test (Raven, Court & Raven, 1995; Richardson, 1991). The data generated were treated

statistically and analysed to provide descriptive statistics. To account for the fact that the different groups did not start at exactly the same baseline when pre-tests were conducted, analyses of co-variance (ANCOVA) were applied with pre-test scores being the co-variates to compare the adjusted post-scores. Cronbach's alpha coefficient was used as an indicator of test reliability.

Percentile norms for different age groups in the United Kingdom are provided in the Raven's Test handbook, e.g., the United Kingdom percentile norms for 12 year-old children are 15 (5 percentile), 21 (ten percentile) and 38 (50 percentile or median score). The RSPM test consists of five sets (A – E) of 12 problems (60 in all), which are graded in terms of increasing difficulty.

As the learners participating in this study are all Xhosa home-language speakers (but who are taught in English) the tests were administered in English by a fieldworker who is also a Xhosa home-language speaker so that she could answer queries made by the participants in their home language if necessary.

Before any inferences could be made as to possible effect on learner cognition by the project it was important to determine whether discussion had taken place in the classrooms of the experimental schools during the period of intervention. (As noted earlier, the assumption that meaningful discussion seldom, if ever, occurred in these classes was premised on a wide range of South African research data, classroom observations of the participating teachers just prior to the intervention and extensive experience of science teaching in Eastern Cape schools.) It was mainly to this end that the qualitative results were generated. Firstly, a four-point scale classroom observation instrument was developed in order to record the classroom activities that took place

during classroom observation sessions. Each teacher was visited a minimum of three times during the duration of the project – this was deemed sufficient to establish if changes in their practice had occurred as they all were recent past students whose classes had been visited and evaluated by the researcher and fieldworker. Also, on-site discussion with their pupils gave clear indications as to whether teachers were ‘window dressing’ for visits or if their classroom discussion strategies were ongoing. Secondly, study of videotape data on each of the experimental and control groups and analyses of classroom observation records provided insights into the types of discourse and interactions that took place. The criteria used to determine whether classroom discussion had taken place were the ability of learners to engage in the lexicon (use the words appropriately), use scientific explanations (apply connectives) and engage in discourses that included descriptions, predictions, explanations and arguments. While a minimum criterion was used as a ‘cut-off’ point for judging whether classroom discussion had taken place or not (namely that each of the above interactions had been exhibited at least once, and that two of the three were exhibited three or more times per classroom observation), qualitative evaluations of the quality of interactions were also important when determining whether classroom discussion had taken place or not.

All of the videos, narratives and interview reports were reviewed by the first author and the fieldworker. Together, they made a collective judgement as to whether discussion had taken place in any particular lesson or not, eliminating the need to establish any numerical inter-observer reliability measure.

Apart from unstructured interviews with teachers at the end of each classroom observation session, each experimental group teacher was interviewed at the end of the

implementation phase of the classroom discussion initiative. Questions they were asked included: How many science lessons were taught during the period of the intervention? Did you apply the 'classroom discussion' principles to science lessons during the period of the intervention? Did you apply the 'classroom discussion' principles to other lessons during the period of the intervention? How much time did you spend on 'classroom discussion issues during this period? Do you think you are beginning to generate authentic discussion in your classroom? How did you respond to bilingual issues in your classroom? Do you feel that the initiative has made any difference to teaching and learning in your classroom?

These data were then used to inform the design of the second study, which was a repeat of the first study, except that new sets of teachers participated and all three sets of trigger materials were used in the experimental classes to generate classroom discussion.

Qualitative data analysis

Narratives of classroom activities and ensuing discussion were generated by lesson observation, video-recordings of lessons, field notes and interviews with teachers. These narratives were then used to determine whether discussion that met the criteria used in this study, i.e. that learners engaged in the lexicon (used the words appropriately), used scientific explanations (apply connectives) and engaged in discourses that included descriptions, predictions, explanations and arguments, had taken place.

Time scales

Both the first and second studies ran over a period of six months, starting in the first quarters of 2001 and 2002 respectively (the starting dates of each grade in schools in South Africa). The participating teachers had been identified and observed in their classrooms during the last quarter of the previous year and the one-day workshops on classroom discussion and the use of the trigger materials took place during the second and third weeks of the first term of the new grade year. The Raven's Standard Progressive Matrices pre-test was administered during the third and fourth weeks of the term. The four topics in magnetism then were taught over a period of eight weeks, which took the classroom intervention phase of the project into the second quarter of the year. The Raven's post-test was administered within two weeks of completing the classroom intervention phase of the project and post-implementation interviews were held with the participating teachers before the term ended.

RESULTS

The results are presented in response to the two guiding research questions.

Discussion

Qualitative data from the classroom observations in the first study revealed that discussion did take place in the majority, but not all, of the experimental groups' post-intervention lessons observed in terms of the criteria for discussion used in this study. It also revealed that discussion of an exploratory nature had not taken place in any of the

comparison groups' lessons. The second study findings on the discussion aspect of the project were similar to those of the first study, i.e. that discussion did take place in the majority of 'treatment' classrooms and that no exploratory talk took place in the comparison classrooms.

Interviews with each of the teachers in the experimental group revealed that they considered that they had applied the principles of classroom discussion to all of the science lessons that they taught from when they had been introduced to the strategy until the time of the interview at the end of the project. Also, these teachers, believed that they had applied the principles, whenever possible, in all of the other subjects that they taught. However, one teacher noted that she believed that it was difficult to apply classroom discussion techniques to mathematics classes. Another teacher noted that she did not enjoy the initiative to begin with and felt frustrated as she "did not know what was expected", but that things had become clearer through contact with the fieldworker after classroom observations and that, in the end she "really enjoyed" the project.

All members of the experimental group of teachers felt that they were beginning to generate authentic classroom discussion (in terms of the criteria made explicit in their training workshops) and that this added a new dimension for their pupils. All agreed that making the 'rules of the game' of classroom discussion clear helped a great deal in terms of having their pupils participate. However, the greatest impact on the teachers seemed to have been made in the area of code switching (switching between the language of instruction and the pupils' home language) and explicitly recognising and valuing mother tongue dialogue. The overall perception was that this was a major factor in enabling their pupils' to engage in meaningful discussion. When interviewed, the comparison teachers

confirmed that they had not used, or known of, any specific classroom discussion strategies, nor that discussion promoting techniques were being used by their peers in the other groups.

Raven's Test Scores

In the first-study, only the urban groups were post-tested. Data from the Raven's Standard Progressive Matrices pre- and post-tests of reasoning (60 items) generated by the first study (n = 146) showed that the frequency distribution of the pre-test was wide and multi-modal, there was a substantial fraction of very low scores and the average score was low. The mean pre- and post-test scores of the four groups are shown in Table 1.

[Place Table 1 about here]

Based on the analysis of covariance (ANCOVA) results of the Raven's data, there is no difference between the four groups at the 99% level of confidence ($p < 0.01$). The 50 percentile score for the first study pre-test fell at a value of 21, while the United Kingdom (UK) 50 percentile norm for 12 year-old children falls at a value of 38. However, the post-test 50 percentile for the Port Elizabeth (urban) sample (including the control group) fell at 35, a figure that approximates that of the UK 50 percentile norm and which represents a considerable improvement in scores.

There was a statistically significant difference between the change in Raven's Standard Progressive Matrices (RSPM) scores of both the practical activities group and the conversational readings group versus the comparison group ($p < 0.05$ in both cases). There was no statistically significant difference between the change in scores of the

practical activity group and the conversational reading group ($p=0.76$), but they both differ significantly from the change in scores of the prompt poster group ($p=0.028$ and 0.01 , respectively). No statistically significant difference was found between the scores of the prompt poster group and the comparison group data ($p=0.69$). For both the practical and conversational readings groups, the effect size was more than one half of one standard deviation suggesting that the magnitude of the pre-post changes were moderate and educationally important. Cronbach's coefficient alpha (α) was 0.88 , indicating that the data generated was reliable.

Data from the Raven's Standard Progressive Matrices (RSPM) second-study pre- and post-tests of reasoning ($n=1192$; $\alpha = 0.84$) indicated a statistically significant difference ($p=0.000$) between the pre- and post-tests scores of the experimental group, and between the change in score of the experimental group compared to the comparison group. The five sets of 12 problems (60 in total) were analysed separately to investigate the levels of complexity of reasoning at which gains were made. The data generated are indicated in Table 2. The group scores (for sets A – E) are summed to generate the total mean values in Table 2, which are lower than, but comparable to, the means shown in Table 1. A possible reason for the higher scores was that, overall, the children in the urban schools used in the first study had a higher level of ability in the English language.

[Place table 2 about here]

Analysis of variance (ANOVA) of the Raven's second study data indicated that the greatest differences between the experimental group's and comparison group's mean gain scores were in sets A and B ($p<0.001$). The effect sizes are moderate, but are high enough to be considered of practical importance, i.e., greater than 0.2 (the lowest effect sizes were recorded in the rural and peri urban groups which were not post-tested in the first study – see Table 3). There was a statistically significant difference at the $p<0.05$ level

between the experimental and comparison groups in set D. No statistically significant difference between the experimental and comparison groups was recorded in sets C and E ($p=0.16$ and 0.84 , respectively).

Different socio-cultural milieus

In order to examine the second question in this study, i.e. whether the different socio-cultural *milieus* in which the schools are set is reflected in the degree of cognitive development that takes place, the scores of the experimental groups were aggregated according to their socio-geographical areas (i.e., urban, peri-urban and rural) are shown in Table 3. The mean and gain scores for the comparison groups are not included in this table as none of the mean score changes were statistically significant.

[Place Table 3 about here]

Statistically significant differences between overall RSPM pre- and post-test scores by pupils in the urban, peri-urban and rural groups of schools were recorded at the $p<0.01$ level. The greatest total gain was measured in the urban group (5.68), followed by the rural group with a gain of 3.83 and then by the peri-urban group with a gain of 3.22.

CONCLUSION

The quantitative data generated by the *Raven's Standard Progressive Matrices* tests is unequivocal that there is a clear and statistically significant improvement in the mean scores of pupils who participated in classroom discussion initiatives over those of the comparison groups. This result was consistent in both the first and second studies.

Despite the failure of the prompt-poster trigger to result in statistically significant increases in Raven's scores over those of the comparison group in the first study, qualitative data in both the first study and second studies suggest that there were no real differences between the different triggers in terms of initiating discussion. Classroom observations of the teachers who used the prompt-posters in the rural and peri-urban areas (where post-testing did not take place in the first study), and of the teachers who used the prompt-posters in the second study, indicated that this trigger initiated classroom discussion of a similar type to the two other triggers. The particular teacher who used the prompt-posters in the first study (and who was not able to generate discussion as intended) appeared to struggle with the magnetism concepts herself and was nervous. This may have contributed to her inability to generate meaningful classroom discussion in any of the lessons observed.

Generally, the Raven's test results appear unambiguous and easy to explain. The most significant improvements made were in question sets A and B, the least challenging of the progressively more difficult sets of questions. However, the statistically significant improvement in question set D requires some explanation. The explanation offered is framed in terms of the multi-modality of the Raven's data (the explanation of which, in turn, is made within Raven's description of mental development in childhood as 'more like salmon leaps in the stream of life than the equally arranged rungs of a ladder' (Raven *et al.*, 1995). This finding suggests that even though the mean initial (pre-test) scores were low, there were modal groups of pupils who were already scoring fairly well on question sets A and B and who were able to improve on their post-test scores in question set D as a result of intellectual stimulation. As the scores in question set D started from a

very low base (pre-test data), an improvement in scores by a relatively small number of participants could produce a statistically significant result. Also, as with improvements recorded in question sets A and B, this could be attributed, in part at least, to the ‘environmental influences and cultural opportunities’ alluded to by Raven et al. (1995) - in this case taken to be exposure to classroom discussion activities.

Explanations for the improvement between the Raven’s pre- and post-test scores by the comparison groups (which were not statistically significant) could possibly be that the teachers in the comparison group made use of the materials in such a way that their children did interact more than they had previously, although not in a way that can be described as exploratory talk, or that the improvement were a result of the ‘Hawthorne Effect’ and practice - the issue of improved scores caused by having done the test once before is acknowledged by Raven et al. (1995). However, it is the fact that the changes in scores are statistically significantly lower than the scores of the experimental groups is what is of importance.

There was a significant difference between the gains in Raven’s test scores of the three socio-geographical groups. It is difficult to venture an explanation for the fact that the urban schools scored the highest mean gain as it would be expected that any advantages that they may accrue from an urban environment would also be reflected in terms of them having the highest baseline score (the peri-urban group had the highest scores in both the first and second study pre-tests). Similarly, the fact that the rural schools had the second highest gains could be attributed to the low base from which they started, but we still do not have any clear explanation as to causes of the differential gains. The effects of ‘social capital’, local environments and initial language capability

are complex factors that need more careful consideration than was possible in this study, and which merit further research.

The significance of this study in terms of the work already published by Mercer et al. (1999) and Wegerif et al. (1999) is that their focus was on student's ability to solve the Raven's tests of graphical puzzles after exploratory talk which focused on how to solve these particular tests, an approach that could lead to the charge that exploratory talk in this context was akin to 'teaching to the test'. However, in this study the exploratory talk was focused on an aspect of the science curriculum that was in no way directly linked to solving Raven's tests, but which produced similar results to those obtained by the researchers mentioned above.

REFERENCES

- ANC (African National Congress). (1994). *A policy framework for education and training*. Johannesburg: ANC.
- Barnes, D. & Todd, F. (1977). *Communication and learning in small groups*. London: Routledge and Kegan Paul.
- Barnes, D. & Todd, F. (1978). *Discussion and learning in small groups*. London: Routledge and Kegan Paul.
- Barnes, D. & Todd, F. (1995). *Communication and learning revisited*. Portsmouth, UK: Heinemann.
- Bereiter, C., & Scardamalia, M. (1979). Pascual-Leone's M construct as a link between cognitive-developmental and psychometric concepts of intelligence. *Intelligence*, 3, 41-63.

- Brown, D. (1998). *Educational policy and the choice of language in linguistically complex South African schools*. Durban: Education Policy Unit, University of Natal.
- Bruner, J. (1990). *Acts of meaning*. London: Harvard University Press.
- Carpenter, P., Just, M. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404-431.
- Chisholm, L. (1993). Education Policy in South Africa. Speech presented to South African Society of Education. Soweto. 6 February 1993.
- Christie, F. & Martin, J. (1997). *Genre and institutions: Social practices in the workplace and school*. London: Cassel.
- Crouch, L. & Mabogoane, T. (1997). *Key numerical indicators 1991-1996*. Johannesburg: Centre for Education Policy Development, Evaluation and Management.
- de Lemos, M.M. (1989). *Standard Progressive Matrices, Australian Manual*. Hawthorne, Victoria: The Australian Education Council for Research.
- Department of Education. (2002). *Revised National Curriculum Statement Grades R – 9*. Pretoria, South Africa.
- Dillon, J.T. (1994). *Using discussions in the classroom*. Buckingham, UK: Open University Press.
- Dix, K.L. (1998). *Enhanced Mathematics Learning? Technology-rich versus Traditional Methodology* Unpublished B Ed Honours Thesis, School of Education, Flinders University, Adelaide

- Edwards, D. & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Methuen.
- Enslin, P. (1990). Science and doctrine: Theoretical discourse in South African teacher education. In. M. Nkomo (Ed.). *Pedagogy of domination: Towards a democratic education in South Africa*. New Jersey: African World Press.
- Flanders, N. (1970). *Analysing teacher behaviour*. London: Addison-Wesley.
- Galton, M. & Williamson, J. (1992). *Group work in the primary classroom*. London: Routledge.
- Gardner, S. (1995). Inquiry is no mere conversation. *Critical and Creative Thinking*, 3 (2), 38-49.
- Hartshorne, K. (1992). *Crisis and challenge: Black education 1910-1990*. Cape Town: Oxford University Press.
- Hofmeyr, J. (1993). *Policy change in teacher development*. Mimeographed.
- Lemke, J.L. (1990). *Talking science: Language, learning and values*. New Jersey: Ablex.
- Leont'ev, A.N. (1981). *Problems of the development of mind*. Moscow: Progress Publishers.
- Lipman, M., Sharp, A. M. & Oscanyan, F.S. (1980). *Philosophy in the classroom*. Philadelphia: Temple University Press.
- Macdonald, C.A. (1990). *School-based learning experience*. Pretoria, South Africa: Human Sciences Research Council.
- Macdonald, C.A. (1991). *Eager to talk, learn and think: Bilingual primary education in South Africa*. Cape Town: Maskew Miller Longman.

- Mercer, N. (1996a). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction*, 6 (4), 359-375.
- Mercer, N. (1996b). Sociocultural perspectives and the study of classroom discourse. In C. Coll and I. Edwards (Eds.), *Discourse and learning in the classroom*. Madrid: Infancia and Aprendizaje.
- Mercer, N., Wegerif, R. & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, 25 (1), 95–111.
- Mortimer, E., & Scott, P. (2000). Analysing discourse in the science classroom. In R. Millar, J. Leach & J. Osborne (Eds.), *Improving science education: The contribution of research* (pp. 125-142). Buckingham UK; Philadelphia, PA: Open University Press.
- NEPI (National Education Policy Investigation) (1992). *Teacher education*. Cape Town: NECC/Oxford University Press.
- Newman, D. Griffin, P. & Cole, M. (1989). *The construction zone: Working for cognitive change in school*. Cambridge: University Press.
- Raven, J., Court, J. & Raven, J.C. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologists Press.
- Resnick, I., Salmon, M., Zeitz, C., Wathen, S., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11 (3 & 4), 347-361.
- Richardson, K. (1991). Reasoning with Raven in and out of context. *British Journal of Educational Psychology*, 61 (2), 129-138.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford: Oxford University Press.

- Rogoff, B., Gauvain G. & Ellis, C. (1991). Development viewed in its cultural context. In P. Light, A. Sheldon and M. Woodhead. (Eds.). *Learning to think*. London: Routledge.
- Rojas-Drummond, S., Hernandez, G., Velez, M & Villagran, G. (1998). Cooperative learning and the appropriation of procedural knowledge by primary school children. *Learning and Instruction*, 8 (1), 37-63.
- Rodseth, V. (1995). *Bilingualism and multilingualism in education*. Johannesburg: Centre for Continuing Development.
- Sheeran, Y. & Barnes, D. (1991). *School writing: Discovering the ground rules*. Milton Keynes: Open University Press.
- Sinclair, M. Mc.H & Coulthard, R.M. (1975). *Towards an analysis of discourse. The English used by pupils and teachers*. London: Oxford University Press.
- Solomon, J. (1994). The rise and fall of constructivism. *Studies in Science Education*, 23, 1-19.
- Sprod, T. (1995). Cognitive development, philosophy and children's literature. *Early Child Development and Care*, 107, 23-33.
- Sprod, T. (1997). 'Nobody really knows': The structure and analysis of social constructivist whole class discussion. *International Journal of Science Education*, 19 (8), 911-924
- Taylor, N. & Vinjevold, P. (1999). *Getting learning right*. Report of the President's Education Initiative Research Project. Johannesburg: Joint Education Trust.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge MA: MIT Press.

Vygotsky, L.S. (1981). The genesis of higher mental functions. In: J.V Wertsch (Ed), *The concepts of activity in Soviet psychology*. New York: Sharpe.

Wegerif, R. Mercer, N, & 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, 493-516.

Wertsch, J.V. (1991). *Voices of the mind*. New York: Harvester.

Table 1: Mean pre- and post-test scores, standard deviations, gain scores and effect sizes for the Raven’s test for the four trigger situations in the first study urban group

Group	n	Mean		Standard Deviation		t-value	Effect Size
		Pre	Post	Pre	Post		
Practical Work *	30	24.2	30.87	12.9	12.38		0.53
Prompt Posters	33	23.88	26.76	11.12	11.2		0.26
Big Book *	35	27.43	33.69	9.31	9.06		0.68
Comparison	48	28.37	30.58	12.22	10.28		0.20

N = 146, Cronbach α = 0.88

* p<0.05 level

Table 2: Mean pre- and post-test scores, standard deviations, gain scores and effect sizes for each question set of the Raven’s test for the experimental and comparison groups

Raven Problem sets	Group	Mean		SD		Gain	Effect Size
		Pre	Post	Pre	Post		
A **	Experimental	7.76	8.50	2.87	3.10	0.74	0.25
	Comparison	8.56	9.12	2.44	2.95	0.56	0.21
B **	Experimental	5.00	6.55	3.08	3.46	1.55	0.47
	Comparison	5.79	7.15	2.69	2.97	1.36	0.48
C	Experimental	3.51	4.40	2.54	2.86	0.89	0.33
	Comparison	3.82	4.61	2.21	2.65	0.79	0.32
D *	Experimental	3.88	4.83	2.89	3.09	0.95	0.32
	Comparison	4.22	4.57	2.51	2.63	0.35	0.14
E	Experimental	1.09	1.23	1.13	1.23	0.14	0.12
	Comparison	0.98	0.99	1.02	1.13	0.01	0.01
Tot *	Experimental	21.3	25.5	10.12	11.43	4.26	0.39
	Comparison	23.3	26.5	8.52	9.81	3.07	0.33

N = 1192, Cronbach α = 0.84

* p<0.05 level; ** p<0.01 level

Table 3: Mean pre- and post-test scores, standard deviations, gain scores and effect sizes for each question set of the Raven’s test for the experimental groups in three socio-geographic areas.

Raven	Centre	Pre-Test			Post-Test			Gain	Effect Size
		N	Mean	S.D.	N	Mean	S.D.		
A	Urban	161	7.92	2.59	168	9.02	2.67	1.10	0.42
B	Urban	161	5.12	3.30	168	6.82	3.50	1.70	0.50
C	Urban	161	3.54	2.77	168	4.92	3.04	1.38	0.47
D	Urban	161	3.72	3.14	168	5.14	3.38	1.42	0.43
E	Urban	161	1.21	1.21	168	1.30	1.36	0.09	0.07
Tot	Urban	161	21.52	10.95	168	27.20	11.86	5.68	0.50
A	Peri	130	9.26	1.85	123	9.50	2.17	0.24	0.12
B	Peri	130	6.32	2.80	123	7.91	2.97	1.59	0.55
C	Peri	130	4.62	2.57	123	5.12	2.64	0.50	0.19
D	Peri	130	5.22	2.66	123	6.04	2.58	0.82	0.31
E	Peri	130	1.18	1.08	123	1.24	1.13	0.06	0.05
Tot	Peri	130	26.60	8.18	123	29.82	8.91	3.22	0.38
A	Rural	174	6.49	3.17	174	7.29	3.63	0.80	0.23
B	Rural	174	3.91	2.66	174	5.34	3.34	1.43	0.47
C	Rural	174	2.67	1.91	174	3.39	2.53	0.72	0.32
D	Rural	174	3.02	2.44	174	3.68	2.74	0.66	0.25
E	Rural	174	0.91	1.08	174	1.14	1.18	0.23	0.20
Tot	Rural	174	17.00	8.60	174	20.83	10.99	3.83	0.39

N = 930, $\alpha = 0.88$