*Cambridge Journal of Education*
Vol. 39, No. 1, March 2009, 119–140

# Group work and whole-class teaching with 11- to 14-year-olds compared

Maurice Galton[a]*, Linda Hargreaves and Tony Pell[b]

[a]*Faculty of Education, University of Cambridge;* [b]*School of Education, University of Leicester*

*(Received 12 June 2008; final version received 18 December 2008)*

This article compares the academic performance and classroom behaviour of pupils when taught new concepts or engaged in problem solving in sessions organised either as cooperative group work or whole class, teacher directed instruction. Comparisons of attainment were made in classes of pupils aged 11 to 14 years (Key Stage 3) in English, mathematics and science. Pupils were also observed, mainly during the introductory phase of the topic under investigation, using a specially designed structured observation schedule. The attainment results suggest that a grouping approach is as effective, and in some cases more effective, than when whole class teaching is used. Classroom observation indicated that there were more sustained, higher cognitive level interactions when pupils worked in groups than during whole class discussions. It is argued in conclusion that the group work results could be improved still further if teachers gave more attention to training pupils to work in groups and if more time was given to debriefing after group work.

**Keywords:** grouping; classrooms; attainment; secondary education; observation

## Introduction

The debate about the success or failure of the British Labour government's attempted reforms of literacy and numeracy teaching in England has tended to concentrate, for the most part, on the outcomes at primary level (Brehony, 2005) but proposed changes for pupils aged 11 to 14 years (Key Stage 3) have not been without their critics (Stobart & Stoll, 2005; Tomlinson, 2005). Experienced teachers have also attributed the decline in secondary school pupils' behaviour as identified by the schools' inspectorate (Ofsted, 2005) mainly to the demands of the 'Key Stage 3 National Strategy' (DfES, 2003), and the accompanying high-stakes testing and target setting (MacBeath & Galton, 2004). The pressure to respond to these demands, it was argued, has forced teachers to limit the amount of pupils' active participation in lessons in favour of whole-class direct instruction. Contrary to pupils' expectations that, 'things would be different after transfer to secondary school', the experiences of their first year in their new school (Year 7) have tended to replicate those of the final year of primary school (Year 6), where high-stakes testing is even more likely to restrict curriculum opportunities. In a more recent study of secondary teachers' views about their work lives, in which the same schools were revisited, classroom practitioners have argued that the problems of disaffection among 11- to 14-year-old pupils has increased in the intervening years (Galton & MacBeath, 2008).

*Corresponding author. Email: mg266@cam.ac.uk

ISSN 0305-764X print/ISSN 1469-3577 online

DOI: 10.1080/03057640802701994
http://www.informaworld.com

In this paper we report findings from the Social Pedagogical Research in Group Work (SPRinG) [1] project, which, as explained in the introduction to the issue, was part of the British Economic and Social Research Council's (ESRC) Teaching and Learning Research Programme (TLRP, 2008). The SPRinG project investigated cooperative group work in three age phases: 5 to 7 years (Key Stage 1), 8 to 11 years (Key Stage 2) and 11 to 14 years (Key Stage 3). Its aim was to improve learner outcomes in attainment, attitudes and interactive/dialogic behaviour, through the increased use of group work. This paper is concerned specifically with Key Stage 3 pupils' attainment and interactive behaviour in English, mathematics and science lessons, taught either in group work or whole class organisational settings. Given the persistent gender gap notably in English attainment in this age group (DfES, 2007; Younger & Warrington, with McLellan, 2005), the effects of these settings on boys' and girls' results were investigated also. We return now to an overview of relevant policy initiatives and research to provide a context for the study, and in particular, to consider reasons for teachers' seeming reluctance to use cooperative groups in their teaching.

In developing the Key Stage 3 National Strategy the then DfES (Department for Education and Skills) produced training and guidance materials concerning the use of group work in the teaching of numeracy and literacy (DfES, 2004). However, in the pilot study of the Key Stage 3 Strategy, Stoll *et al.* (2003) found that whole-class teaching continued to dominate the pedagogy of the core subjects, although when pupils were asked about their preferred way of learning 90% of the sample of Year 8 (12- to 13-year-old) pupils expressed a preference for working in groups rather than for whole-class teaching. This finding was supported in a study by Kutnick, Blatchford, Clark, MacIntyre, and Baines (2005a), who surveyed a mix of 250 classes of 12- and 15-year-old pupils from 47 secondary schools. The result of this 'mapping' survey showed that the pupil groupings in these classrooms were rarely determined by the nature of the task demand. When pupils worked in groups they were often allowed to choose their composition, which because they were often based on friendship tended to be strongly associated with the attainment level, gender and ethnicity of the pupils, although this last variable was not included in Kutnick *et al.*'s (2005a) analysis, possibly because of the uneven distribution of minority ethnic groupings across the sampled schools. In science, the determining factor which decided whether group work would be carried out was the availability of equipment. In other cases the decision to move to group work was often arbitrary. For example, in one instance reported in Kutnick *et al.* (2007) the teacher's rationale for changing from 'class' to 'groups' was 'because children were becoming bored with the class work and I thought they needed a change'.

## Possible impediments to the use of group work

The seeming reluctance on the part of secondary teachers in England to use groups specifically to facilitate desired academic outcomes rather than merely for social, practical, or convenience motives, is surprising given the evidence that elsewhere, particularly in the US and Israel, there have been a host of comparative studies setting the use of cooperative and collaborative groupings against more traditional classroom structures. Group work generally results in academic gains as well as social and attitudinal improvements. A number of meta-analyses (Kulik & Kulik,

1992; Lou *et al*., 1996; Slavin, 1983, 1987) have produced moderate effect sizes in favour of groups, although both social and academic gains can be considerable when the tasks are structured in ways that facilitate learning at a metacognitive rather than procedural level (Cohen, 1994). While the academic gains may be limited they are usually accompanied by much greater positive changes in attitude. Given the expressed concerns of teachers about the decline in the behaviour of pupils, it might be thought that the use of groups to encourage cooperative learning might have become more frequent since the introduction of the Key Stage 3 Strategy.

There may be a number of reasons why this transformation in teaching methodology has not occurred, in addition to the one cited earlier that teachers are under too much pressure, because of the demands of the curriculum, to find the necessary time for setting up and carrying out group tasks. Other practitioners may have concerns about loss of control over the learning environment since it is often difficult to be certain that the talk among pupils is productive, or even on-task. This contrasts sharply with the situation during whole-class discussion where pupils generally only speak when requested to do so by the teacher. Teachers may also see the need for mixed ability groups as an additional source of difficulty, since it often means placing a disruptive pupil among others who work well together with possible problematic consequences. However, Putnam, Markovchick, Johnson, and Johnson (1996) found that when such children are placed with others there are positive socialising effects. By way of contrast, Brinton, Fujiki, and Higbee (1998) suggests that aggression or withdrawal is a more typical response. There is also some evidence that suggests that different personality types react differently to the groups. Webb (1989) for example found that introverts were more likely to withdraw from groups and in Kutnick *et al*'s (2007) case studies many teachers said that they did not know what to do with the 'loners' who withdrew from group activity completely. Entwistle (1977) also suggests that anxious pupils learn best when materials are highly structured as in whole-class direct instruction so that the more open-ended problem solving' approach which researchers such as Sharan and Sharan (1992) claim facilitates cooperation within groups may not be so appropriate for such pupils.

Another key reason for teachers' reluctance to use grouping to enhance academic performance may be the requirement that to work effectively pupils need to be trained in communication skills (Mercer, 2000). In the US, Webb and colleagues (Webb, Baxter, & Thompson, 1997; Webb & Farivar, 1994; Webb & Mastergeorge, 2003) have studied the communication processes necessary to ensure effective group work. They emphasise the necessity of having pupils in the groups who can undertake focused questioning, explore alternative answers and provide explanations for these answers if groups are to be effective in their problem-solving. Pupils who provide such help have been shown to benefit academically as a result of facilitating this 'high level' cognitive discourse, as do those to whom the help is given (Webb, 1991). To be able to operate at this cognitive level, however, requires appropriate training, not only to enhance this exploratory talk but to develop the necessary levels of trust among group members so that in subjects such as humanities and English they can handle potentially controversial or sensitive issues such as race, gender or politics. Kutnick *et al*.'s (2005b) review of secondary school teachers' approach to group work, found that teachers provided little training for either social or communicative interactions.

Given this gap between what research recommends and what classroom practitioners in secondary schools typically do, it was decided to include a Key Stage 3 element in the SPRinG project. The relative dearth of UK studies at secondary level, a notable exception being that of Cowie, Smith, Bolton, and Lever (1994) and Cowie and Rudduck (1988) also favoured an attempt to replicate more recent American findings but, in keeping with the principles of social pedagogy, in naturalistic classroom settings. Having thus set out the background to this research, the remaining sections of the paper will describe the methodology employed, then present the main results and finally discuss the implication of these findings for both future research and current practice.

## Design of the study

### *Participants*

As explained in the introduction to this paper, this third strand of the SPRinG project was conducted with 11- to 14-year-old pupils in Key Stage 3 classes. Initially, 14 English and 16 science teachers were recruited but in the second year of the study a further cohort of 12 mathematics specialists were added. The sample consisted mainly of volunteers. In the first instance all schools in the Cambridge Initial Teacher Training Partnership at secondary level were written to and invited to participate. Other teachers were added to these numbers by asking the senior advisor of one neighbouring local education authority to nominate further participants. The advisor was asked to provide a range of practitioners, among whom there would be some already strongly committed to the use of cooperative learning within groups.

### *Design*

Originally it was hoped to conduct a 'true' experiment with teachers randomly assigned to two conditions (use of group work versus whole-class and individual seat work) but this proved not to be feasible for a number of reasons. First, and most importantly, few teachers were willing to be part of the whole-class and individual seat work treatment group. All said that they had joined the project in order to discover how to use group work more effectively and that they therefore did not wish to be excluded from participating in the various training programmes. Secondly, teachers argued that in some cases, particularly in science where for some experiments there were limited amounts of equipment, bringing pupils together in groups was the only feasible strategy.

After some discussion with the participating teachers it was agreed that the key difference between the use of groups for, say, practical tasks and for promoting cognitive development was in the use of groups for developing new concepts or for problem solving. Typically, the former was usually done through class discussion involving what Edwards and Mercer (1987) have described as 'cued elicitations'. It was agreed, therefore, that the essential comparison would be that in all cases where it was intended to introduce new concepts, develop ideas and find solutions to problems etc, then this would be done either by means of whole-class activity or through the use of small groups. At other points in time where the tasks involved were practical or relatively low level, as in mathematics where pupils might be comparing answers or solving different aspects of a problem individually and then bringing the final result

together, then those teachers allocated to the whole class approach could also use groups for these latter purposes. The essential distinction here was between the three different types of grouping as defined by Galton and Williamson (1992). The first kind of grouping was designated collaborative. In this situation there was complete social inter-dependence and shared accountability so that pupils worked collectively on the same activity to complete a common task. For example, in English, pupils might have to make predictions about what certain characters in a story might do next, while in science, faced with certain results, they might be asked to come up jointly with an acceptable hypothesis and way of testing it.

The second kind of grouping was termed cooperative. Here pupils have individual accountability as they work independently on different activities, each of which makes a contribution to a common goal. In mathematics, for example, each pupil having worked out a number of examples might then share these results with the rest of the group, while in science pupils might take measurements of a chemical process at different temperatures and then come together to produce the overall graph showing the rate of reaction.

Finally, pupils could be placed in '*seated*' groups where individuals worked on their own but helped each other by checking and comparing answers. In English this might involve pupils reading each other's stories while in mathematics it would involve them checking each other's calculations. Essentially in this study, comparison was between collaborative forms of group work and whole class discussion for the promotion of higher order level thinking.

In accordance with the principles of social pedagogy the study was designed to follow the normal pattern of events as closely as possible. In mathematics and science, for example, units of work usually lasted for around six to eight weeks. In each subject teachers were randomly assigned to one of two groups. The first group was then randomly assigned to two of four chosen units. The teacher in this group then taught those two units using collaborative group work and the remaining two units using whole-class teaching coupled with individual seat work. The situation was then automatically reversed for the second group of teachers so that whenever the first group taught a unit using collaborative group work, the second would use whole-class teaching. This created a repeated measures design in which each teacher was observed once during the teaching of each topic, using either group work or whole class teaching according to the random assignment described above.

Sixty-nine English lessons were observed in Year 1 of the study. A further 53 lessons, a mix of mathematics and science, were observed as part of the observation schedule development. In the second year 96 lessons in all (again a mix of English, mathematics and science) were observed.

### *The assessed tasks*

Table 1 shows the topics chosen for each subject. In each case the topics were selected by the teachers to fit in with the organisation of their existing schemes of work for each particular year group. In mathematics the topics consisted of number patterns, interpreting data, ratio and areas and volumes. In science the topics involved the study of particles, electric circuits, forces and living cells. English presented more difficulties in that often the goal of the teaching is for pupils to gain an understanding of certain general processes and to acquire certain skills rather than,

Table 1. Measures used to assess pupils' performance.

| English | Science | Mathematics |
|---|---|---|
| 1. Narrative Writing | 1. Particles | 1. Number patterns |
| 2. Response to Text | 2. Electric circuits | 2. Interpreting data |
| 3. Persuasive writing | 3. Forces | 3. Ratio |
| 4. Descriptive writing | 4. Living Cells | 4. Areas & volumes |

as in mathematics and science, where it is usually to gain knowledge and understanding of specific content. It was decided therefore to focus on four forms of writing genre (narrative, response to text, persuasive and descriptive). Teachers of English were free to choose their own content but the end product would be a piece of work which could be assessed using appropriate criteria.

In each case pupils were given a test prior to beginning a unit and the same test was administered once the unit had been completed. Most units typically took six weeks. The science attainment measures were mainly taken from a series of diagnostic tests constructed as part of another TLRP Project (Towards Evidence-Based Practice in Science Education). Where items covering the particular topic were not available others were selected from the Key Stage 3 statutory tests and from the optional tests created by the Qualifications and Curriculum Authority (QCA). In mathematics the number patterns for algebra were largely taken from the *Dime* (Diversity in Mathematic Education) set of mathematics enrichment resources published by Tarquin Press. Items were generally investigative in style testing higher order thinking. Interpreting data for data handling were taken mainly from the Secondary Mathematics Independent Learning Experience (SMILE) pack. Again, there was a mix of items mainly testing higher order thinking. Ratio for number was again constructed from the SMILE pack and from other QCA test examples. These items mainly demanded knowledge application rather than problem solving. Areas and volume for measurement for space and shape were also constructed from the SMILE pack and like the ratio material the items mainly demanded knowledge application rather than problem-solving skills. In the case of both mathematics and science all the items chosen were assigned a level by two expert judges and a mix of items selected which were deemed appropriate for pupils aged 12, 13 and 14 in Years 7, 8 and 9 respectively. The pupils' overall level score was then converted into a number using the NFER numeric conversion scale. This is shown in the appendix.

In English the procedure was somewhat different in that pupils were given an assignment initially involving various kinds of writing. This was marked by a small group of teachers from within the project, all of whom had had experience of public examining at GCSE and were therefore familiar with the process of levelling work. At the end of the six weeks pupils again were asked to complete a similar piece of writing and this was again marked by these experts. In each case the marking was 'blind'. Levels were again converted to numeric scores using the NFER scale.

### *The lesson observation system*

The lessons were observed by trained observers using a structured observation schedule specially designed for the purpose. Table 2 shows the main variables that

Table 2. Categories used to determine pupils' classroom behaviour.

| **Target's Behaviour** | **Target Interaction** | **Target-Adult** |
|---|---|---|
| On-task | Question | Group |
| Routine-wait | Suggests | Class |
| Distracted | Explain/reason | Individual |
| CODS/RIS | Agree | Content |
| Other | Argue/disagree | Task |
| | Seek help | Routine |
| | Maintain | Monitor |
| | Block | Other |
| | Sustains | |

the observer focused upon. The system corresponded to the one originally designed for the ORACLE (Observational Research and Classroom Learning Evaluation) studies and their subsequent follow-up (Galton, Simon, & Croll, 1980; Galton *et al*., 1999). The standard categories on the ORACLE pupil record were used for recording whether pupils were on task, distracted, waiting or engaged in a routine activity, or, in cases of uncertainty, whether they were partially cooperating and partially distracted (CODS), or responding to internal stimuli (RIS). RIS covers situations where pupils might be thought to be daydreaming (e.g., staring out of a window). Various other categories in the ORACLE schedule mainly involving physical or verbal disruption rather than mere distraction or moving around the classroom, and which were very rarely observed in the ORACLE studies, were combined under a general category labelled 'Other'.

A second feature of the ORACLE pupil record was also incorporated. This concerned the audience categories, namely (a) whether the target pupil of the adult's audience was a member of the class, of a group, or being addressed as an individual; and (b) the nature of this interaction; whether the content was about task, a routine matter, whether the adult was simply monitoring the group or class activity or whether some other interaction involving, for example, housekeeping (e.g., collecting books) was taking place. Each set of observations consisted of four consecutive 30 second intervals per target pupil. The observer focused first on the more detailed aspects of interaction described in the next paragraph, and then noted which behaviours belonging to these major categories (target's behaviour and target-adult interaction) were occurring for at least two thirds of each 30 second interval. This procedure differs in this respect from the instantaneous time sampling used in the ORACLE research (Croll, 1986). Furthermore if the target pupil was in continuous interaction with another pupil or the teacher throughout the 30 seconds this was coded as a 'sustained' exchange.

The third feature of the observation schedule was not part of the original ORACLE pupil record. Its purpose was to determine the nature of the target pupils' interaction and the categories were chosen to reflect the findings of earlier studies, particularly those of Webb (1985) and Mercer (2000) concerning the importance of effective communication among group members. In each time unit note was taken as to whether a question was asked, a suggestion made, an explanation or reason given, agreements sought, disagreement expressed. These categories were designed to

capture what might be termed higher level cognitive reasoning. In addition cases were also recorded where pupils in a group sought help or assisted in maintaining the flow of group activity by acting as gatekeepers as against other cases where group members sought to block the discussion through negative contributions. In the case of all these categories, such interactions were recorded once on the first occasion they occurred during each of the four consecutive 30-second time intervals constituting one round of observation per target pupil. This process known as 'one-zero' time sampling is a device occasionally used in systematic observation when the categories of behaviour are infrequent and where their identification is sometimes more difficult to determine reliably. This approach was first used in an earlier study of science teaching (Eggleston, Galton, & Jones, 1976) in order to distinguish between difficult science categories such as hypothesising and making inferences. The use of one-zero sampling has been criticised (Dunkerton, 1981) because it is not possible to claim, as with instantaneous time sampling, that the proportion of the total observations represented by a particular category is an approximate measure of the proportion present in the whole population. In practice, however, in cases where the occurrence of these higher inference categories tends to be comparatively rare and where the time unit is sufficiently short, then such categories are likely only to occur once within any given time unit and can therefore also be said to represent the overall proportion of a given activity within the population. This assumption was adhered to in the present analysis since it was rare that categories involving agreeing or disagreeing, for giving explanations, for example, occurred more than once during a 30 second interval. In the case of those teachers using a whole-class approach, the observation visits generally took place during the introduction to a topic when often ideas were being developed. Where this was not so, then the teacher generally alerted the observer to an appropriate session which could be observed. In the case of the teachers using collaborative group work the observers generally attended lessons where a high level of discussion was planned rather than the lesson consisting of practical activity.

Observer training involved working with video initially in learning to apply the category definitions, lengthy discussions to refine the definitions, paired observations in the classroom, modification of the schedule as necessary to improve validity and inter-observer agreements, and repeated reliability-cum-moderation trials to achieve reliabilities over 0.8 for the principal and more frequently occurring categories.

Finally, as is usual in collecting observational data, the observers recorded contextual information such as classroom layout and lesson details, as well as a post-observation summary of the session including, for example, lesson objectives, details of the task(s) (e.g., practice-abstract), group composition (e.g., size, gender mix), and briefing and debriefing with respect to group work.

### Training the teachers to use group work

Prior to carrying out these observations all teachers had taken part in a programme designed to help them hone their pupils' group working skills. Pupils when working in groups do not necessarily or spontaneously engage in activities that enhance their learning. It has been shown by Kagan (1988) that with training the pupils can improve both the quality and the effectiveness of their work in groups. The SPRinG programme was an amalgam of various approaches including the development of

group dynamics (Kingsley-Mills, McNamara, & Woodward, 1992) the development of suitable attitudes (Farivar & Webb, 1991) and the improvement of various communication skills (Mercer, 2000; Wilkinson & Canter, 1982). In the event, unlike their colleagues in the primary sector who had taken part in the SPRinG project, the secondary teachers found it difficult to give the time to the full training programme. In particular, some of the participating science teachers questioned the necessity of pupils needing to spend time on trust exercises designed to improve the dynamics of groups. As a compromise, it was agreed that training should concentrate on a few selected areas to be completed during the first term of the year. These included the pupils' setting and improving rules governing both their behaviour and conversation in groups, exercises designed to improve group maintenance skills (summarising, time-keeping, etc) exercises designed to improve decision making by consensus rather than by voting, and, at a later stage, exercises designed to support better reasoning and explanation within the groups. Most of these were taken either from Kingsley-Mills *et al*. (1992) or Cowie and Rudduck (1988).

## Results

It proved difficult for teachers to carry out all four tasks during the course of the year. This emerged during the first year of the fieldwork when a trial of the design was carried out among some members of the English teaching cohort. During the construction of the training programme it was found that many of the activities designed to promote better group cohesion were a natural part of pupils' developing response to English text. In many English lessons there was an emphasis on sharing feelings as well as ideas and this required the development of empathetic relationships between pupils during both class and group discussion. English teachers therefore felt ready to undertake the comparative study before either their science or mathematics colleagues. Having found that four writing tasks were too many the English teachers agreed to collapse the categories into two, that of imaginative and discursive writing for the second year of fieldwork. Data collected in the first year on all four forms of writing were then collapsed with response to text and persuasive writing being combined into the imaginative mode and descriptive and narrative writing into the discursive category. Data for both years one and two were then aggregated to give the results shown in Table 3 where the means, standard deviations (in brackets) and effect sizes are presented. It can readily be seen that in the case of imaginative writing the classes taught through the use of collaborative grouping outperformed those in which whole-class teaching was the dominant methodology. The pupils taught mainly through the use of groups raised their performance by nearly a whole level (from 5− to 5+) whereas the scores of those taught as a class barely changed. The result produces a small to medium effect size using Cohen's (1988) criterion which according to Hattie (2005) represents upwards of one-third of a year's progress. For discursive writing those taught in groups again do better than those taught mostly through whole-class teaching. In both cases the gains are significant (at the 1% level for groups and at the 5% level for class). The effect size for groups is nearly twice that for students taught by the whole-class method. Overall, therefore, the findings are very positive in terms of using an approach based on collaborative group work.

In the second half of Table $3^2$ a similar analysis is presented, this time by gender. In the case of boys the gains between pre-test and post-test are significant for both

Table 3. Progress in English by topic and by gender: (s.d. in brackets).

| Topic/Gender | Mode | Pre-test | Post-test | N | Effect-size |
|---|---|---|---|---|---|
| **Imaginative** | Group | 31.17 (5.27) | 32.57** (5.30) | 200 | 0.43 |
| | Class | 29.73 (7.90) | 30.31 (6.50) | 137 | |
| **Discursive** | Group | 29.32 (6.06) | 31.19** (5.34) | 185 | 0.37 |
| | Class | 31.12 (5.61) | 31.80* (5.28) | 178 | 0.18 |
| **Boys** | Group | 29.29 (5.35) | 30.66** (4.93) | 211 | 0.38 |
| | Class | 28.37 (6.19) | 29.13* (5.62) | 186 | 0.18 |
| **Girls** | Group | 30.87 (5.60) | 32.79** (5.38) | 223 | 0.42 |
| | Class | 32.31 (7.17) | 32.84 (6.08) | 211 | |

*=$p< 0.05$; **=$p<0.01$.

groups and class, but the effect sizes in the case of groups is again twice that for pupils taught mainly by a whole-class approach. In the case of girls, however, it is only the approach based on the use of groups that achieves statistical significance, giving a small to medium effect size equivalent to one-third of a year's progress compared to the results for whole-class teaching. On this evidence therefore working in groups to develop pupils' ideas in English has much to recommend it, particularly for boys, although, while more beneficial than whole class work for boys' attainment, it does not close the gender gap.

In dealing with the maths attainment scores a similar procedure was adopted to that used for English, in that the scores on different topics were combined according to the perceived task demand. Scores, based on low level cognitive demand, mainly consisted of items from the ratio and areas and volumes tests while high level scores were mainly taken from the number pattern and data handling items. Here the results, shown in Table 4 are not as clear-cut as were those for English. When the cognitive demand was low level neither pupils taught in groups or mainly through whole-class teaching made significant progress. There is clearly an ability teaching mode interaction in that on the tasks of lower cognitive level the pre-test score indicates that the pupils who worked in groups appear to be of lower ability than those taught by whole-class methods. The situation is reversed when the higher level

Table 4. Progress in mathematics by task demand and by gender (s.d. in brackets).

| Task demand and Gender | Mode | Pre-test | Post-test | N | Effect-size |
|---|---|---|---|---|---|
| **Low Level demand** | Group | 30.35 (4.14) | 30.53 (4.62) | 133 | |
| **(ratio & area/volume)** | Class | 34.64 (4.37) | 35.08 (5.64) | 167 | |
| **High level demand (number** | Group | 33.54 (4.56) | 34.98** (5.22) | 185 | 0.70 |
| **pattern & data handling)** | Class | 29.78 (3.78) | 30.45** (4.26) | 178 | 0.48 |
| **Boys** | Group | 31.16 (4.26) | 32.00** (4.90) | 124 | 0.56 |
| | Class | 32.67 (4.88) | 32.96 (6.15) | 134 | |
| **Girls** | Group | 32.33 (4.91) | 32.90 (5.82) | 108 | |
| | Class | 33.08 (4.69) | 33.87** (5.04) | 127 | 0.54 |

**=$p<0.01$.

scores are considered. This may have come about in one of two ways. Either the data reflect an actual ability difference in that the classes for the low level tasks where group methods were used are mainly of low ability or, and this is the more likely explanation, the difference is a consequence of age so that there were more Year 9 classes in the whole-class teaching cohort and they naturally had higher pre-test scores. This situation then reverses when the higher level cognitive scores are considered. Table 4 shows that when mathematics tasks are concerned with investigations involving higher level thinking both results are significant (at the 1% level) but the effect size achieved in the groups is higher than that achieved through whole-class teaching, although in both cases the gains are sizeable and represent considerable progress. When the scores are looked at in terms of gender there is a further interaction effect in that boys did better when taught in groups (effect size 0.56) while the reverse was true for the girl pupils. The girls made more progress when a whole-class teaching approach was used. There is slight evidence to suggest that this is a genuine effect in that boys taught in groups have the lowest pre-test mean and in a previous paper (Pell *et al*., 2008) it was shown that lower achieving boys have more positive attitudes to group work although the correlation (−0.06) is a very modest one. These positive attitudes to group work are also positively correlated with pupils' anti-learning and anti-school dispositions. These pupils, mainly boys, are precisely those whose mathematics achievement has been the source of so much national concern (Warrington & Younger, 2006).

The science results are more difficult to interpret partly because the decision over which topics to attempt was largely governed by the way the syllabus was constructed and, in particular, the year in which a topic was covered. This tended to vary from school to school. In the event, the decision to offer only two out of the four topics resulted in certain anomalies in that no teacher undertook the living cells module. Because of the different points in time when topics were taught in various schools, most teachers eventually undertook the electricity topic as a whole class activity but had more flexibility in their choice of either particles or forces for the group activity. As a result, there was considerable difference between numbers in the group and class cohorts in each of the three topics. A second problem occurred because in the case of science, the teachers were concerned that the test items, particularly the diagnostic type, covered too large an ability range. It was therefore agreed that each teacher should recommend the items for their class test which they believed matched the average ability level of the pupils that they taught. Initial analysis of the results showed that some teachers had tended to underestimate the ability of their pupils and as a result 'ceiling' effects occurred at the post-test and some pupils achieved the maximum possible score on the particular test used.

The results are shown in Table 5. In all three tests the pupils taught using groups make significant gains with the same magnitude of effect size in the case of electricity (both group and class cohorts achieving effect sizes approaching 0.8, the large Cohen (1988) criterion). However, for both forces and particles there are clear advantages in terms of effects sizes of gains for group working. It appears that working mode is topic dependent. When gender comparisons are examined, there are significant gains for both boys and girls when taught by either group or whole-class methods. In terms of the effect sizes boys do equally well when taught by either method but girls do better when working in groups.

Table 5. Progress in science by topic and by gender (s.d. in brackets).

| Topic/Gender | Mode | Pre-test | Post-test | N | Effect-size |
|---|---|---|---|---|---|
| **Electricty** | Group | 26.85 (4.19) | 30.62** (5.25) | 26 | 0.71 |
| | Class | 29.25 (5.15) | 33.58** (5.49) | 106 | 0.73 |
| **Forces** | Group | 26.21 (7.57) | 29.79** (8.06) | 28 | 0.98 |
| | Class | 31.76 (8.51) | 33.35* (9.74) | 66 | 0.27 |
| **Particles** | Group | 29.32 (5.27) | 31.93** (6.27) | 94 | 0.48 |
| | Class | 31.21 (4.81) | 31.92 (5.15) | 39 | |
| **Boys** | Group | 28.56 (5.46) | 31.30** (6.09) | 54 | 0.53 |
| | Class | 29.87 (7.87) | 33.64** (9.23) | 53 | 0.69 |
| **Girls** | Group | 26.81 (5.86) | 29.89** (5.86) | 53 | 0.58 |
| | Class | 30.00 (7.86) | 33.70** (8.23) | 54 | 0.49 |

*=$p<0.05$; **=$p<0.01$.

## Classroom observation

By working with teachers to induct their pupils into the use of groupwork, the SPRinG project aimed to improve levels of engagement, and the quality of interaction during group work. The classroom observation results (Figures 1 to 4) reveal that there were improvements in both areas, demonstrating simultaneously that the achievement of high level discourse within the groups was more likely to lead to improved learning outcomes. Figures 1 to 4 show how pupils' behaviour and interactions changed during the school year as they and their teachers gained experience in using group work.

Figure 1 shows the relative amounts of time on task made during visits to the classrooms to see either whole class teaching or group work. For each subject one visit took place in the spring and one in the summer term, the class having been trained during the autumn period. For English it can be seen that there is very little difference between those taught in groups and those taught as a whole class in the spring visit but a slight advantage in favour of groups by the summer of around 5%

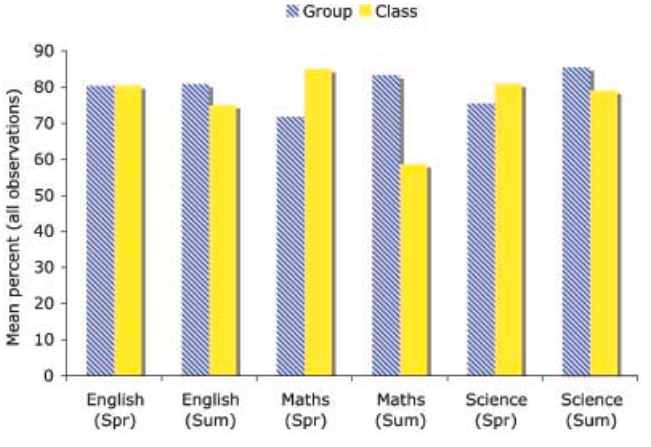


Figure 1. Time on task in group and whole class settings.

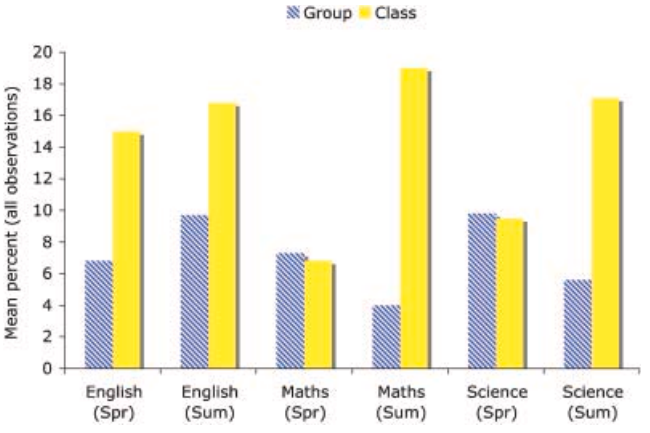


Figure 2. Partial distraction in group and whole class settings.

of all observations. For mathematics the position changes quite dramatically over the two terms. During the spring term, time on task in the groups is around 70% compared to over 80% for whole-class teaching. However, by the summer visits this trend has been dramatically reversed. Time on task in groups is over 80% while the corresponding figure during whole-class teaching has reduced to 58%. Science shows a similar trend, although a less dramatic one. The corresponding figures in the spring term were 76% for groups compared to 81% in class and for the summer term, 84% compared to 79%.

In Figure 2 the corresponding data for the CODS (partially distracted, partially attending) category are given. The data here reflect the fact that it is much harder to determine with complete certainty when children in a whole-class situation are paying attention. As a consequence, the per cent levels of this category are much higher during whole-class teaching than they are in groups. If, however, the assumption is made that the observer's total coding of this category involved approximately equal incidents of pupils who were actually distracted or attending (that is, the observer correctly coded the pupil as being distracted on 50% of the occasions) then the figures still represent a small but significant balance of on-task behaviour in favour of groups, particularly by the summer term.

Figures 3 and 4 show the developments in the pupils' interactions. As pupils become more skilled in group work after training and with practice, so an increase in interactions associated with the kinds of higher level cognitive discourse advocated by Webb (1985) is predicted. These interactions consist of a combination of the *asking questions*, *offering explanations*, *making suggestions*, *agreeing* and *disagreeing* categories. In both English and mathematics in either the spring or summer terms the per-cent figure for the groups are superior. In science, although pupils in the groups outperform those in the whole-class situation the differences are marginal. In interpreting the data in Figure 3 it should be remembered that the process favours the class rather than the group because of the sampling procedures. In the group situation two groups were picked at random and then two pupils in the group were sampled as a target in turn. This procedure was used to code the various categories of pupil behaviour and pupil adult interaction under the time sampling arrangement. But in collecting evidence about the nature of the discourse using the one-zero

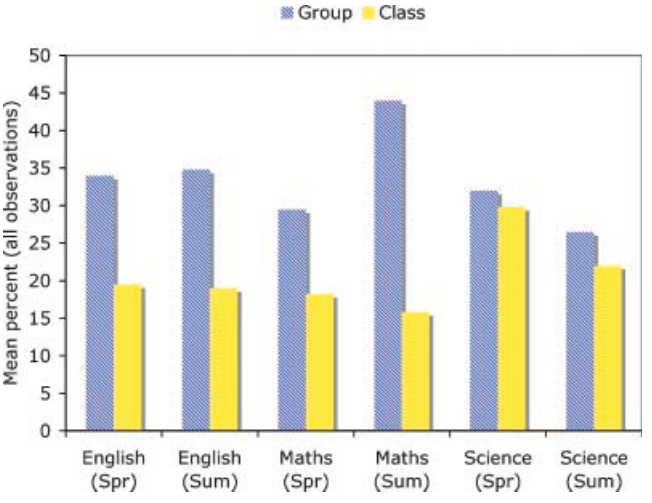


Figure 3. Open dialogue in group and whole class settings.

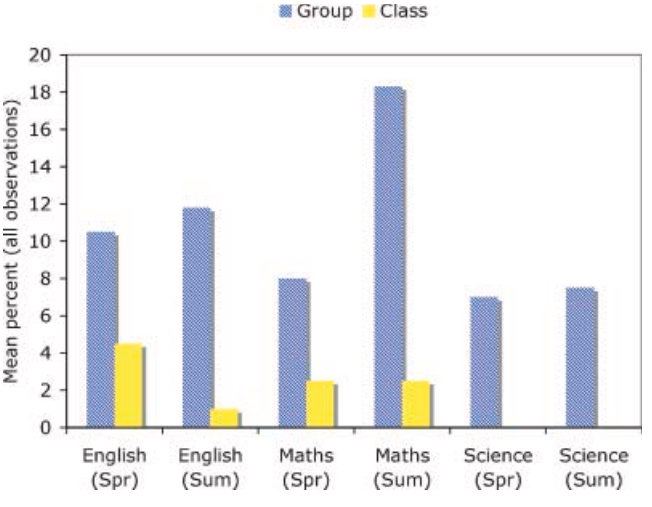


Figure 4. Sustained interactions in group and whole class settings.

sampling procedure any utterance emerging from the group was recorded. Thus in a class consisting of 20 pupils divided into five groups there was a much greater chance of not recording one of the specified utterances because it occurred in a group that was not being targeted whereas in the class situation all pupils were included because they belonged to one group, the class. The data presented here corroborate much of the previous research on whole-class interactive teaching (Alexander, 2006; Hargreaves *et al*., 2003) which typically consists either of extended teacher direction or short sharp question-response sequences (Dillon, 1990). Particularly impressive is the figure for mathematics which exhibits a typical observed profile consisting of rapid closed questions during whole class discussion (Galton *et al*., 1980, 1999).

The final figure (Figure 4) examines the number of interactions which extend over one time unit and were therefore classified as sustained. Again these data tends to confirm previous research on whole-class interactive teaching where there is little extended conversation. Here in English and again particularly in mathematics in the summer term, pupils in groups were able to sustain on-task conversations and it is probable that such extended discourse was a consequence of the pupils' increasing willingness to offer more explanations and to debate and discuss alternatives as represented in the use of these categories in the previous figure. Science presents an extreme case, in that there were no recorded sustained observations during whole-class discussion. Thus we may conclude that, overall, in terms of improved attainment and the quality of interactions taking place, the use of collaborative group work to develop ideas, solve problems, carry out investigations and perform other higher cognitive level activities can be recommended as an alternative to the use of whole-class teaching. Although there are some findings which contradict this view, the trend overall, particularly when the effect on attitudes is also taken into account (Pell, Galton, Steward, Page, & Hargreaves, 2008) tends to favour the increased use of groups in the lower secondary school.

## Discussion and implications

### *Group work and whole class teaching*

This comparative study was not intended to sustain an argument that group work should replace whole class teaching and indeed these findings in themselves do not constitute an overwhelming case for the use of group work rather than whole-class discussion. Rather we argue that group work should be a complementary organisational strategy in promoting conceptual and metacognitive learning and our results support this claim. When the academic outcomes of group work, as demonstrated in this study, are coupled with the wide-spread evidence that use of this teaching approach also promotes better peer relationships, then the case for ensuring that group work no longer remains a 'neglected art' in English classrooms becomes an exceptionally strong one. Thus, just as we must seek to improve the quality of the interactions taking place in groups, so we need to attend also to the present impoverished nature of much whole class discussion.

Alexander (2006) has been working with teachers to improve the quality of interaction in whole class settings as well as between teachers and pupils in individual and group organisations. As he observes, part of the difficulty in implementing the government's initiative to encourage the effective use of what was initially termed 'whole-class interactive teaching' has lain in the failure to define such terms precisely. Thus, in much of the government literature, the advocacy of this approach was often accompanied by an injunction that lessons should maintain a rapid pace. Recommendations of this kind represent a confusion between the use of questioning as in *direct instruction* (Rosenshine, 1979; Gage, 1978) and its use to enable pupils to engage in '*thoughtful discourse*' as a means of teaching for understanding rather than transmission (Good & Brophy, 2002). In direct instruction rapid question and answers are used by the teacher to find out what pupils remember from previous lessons in order to decide whether previous material needs to be re-taught or new material introduced. This is very different from interactive whole-class teaching which is designed, as with group work, to promote thinking and problem-solving

and which therefore requires pupils to have extended *wait times* or as Alexander prefers, '*thinking times*', in order to process and organise information before providing an answer. Alexander (2006) has developed these ideas into a programme which he describes as *dialogic teaching* which in its principles mirrors many of the purposes of group work, although with greater teacher participation. According to Alexander's initial results reading and writing have benefited from the increased emphasis on dialogic talk during classroom discourse particularly for the lower achieving pupils but not all teachers are able to change easily from the dominant use of cued elicitations. It could be argued, therefore, that a truer comparison between whole-class and group instruction should perhaps seek to make comparisons between two treatment groups, one trained in group working skills and one trained in the use of dialogic talk during whole-class teaching sessions. Even so, there are grounds for suggesting that if, in this study, the structure of the group activities had been better organised then the differences in terms of pupil progress between whole-class discussion and talk in groups would have been even larger.

### *Improving group work: debriefing*

There are a number of reasons for claiming that the group work could have been more effective, to do mainly with the context in which teachers had to operate when taking part in this study. First, as part of training pupils to work effectively in groups it is vital that teachers brief and debrief the class so that they can begin to gain metacognitive awareness of what it means to be part of a group. Debriefing sessions therefore are particularly important because they not only evaluate how individuals responded in the groups but they also call for participants to make suggestions about suitable strategies for improving the situation on future occasions. After each session, observers completed a lesson overview schedule which recorded, amongst other things, whether or not briefing or debriefing had taken place. It was noticeable, however, particularly in science, that teachers rarely found time for these debriefing sessions. It was rare, for example, to observe a science lesson where the teacher with, say, five minutes of the period left preferred to keep discussion of the results over until the next lesson and instead engaged in a debriefing exercise. More often teachers preferred to use an evaluation sheet which they handed to pupils as they left the class. Thus the exercise tended to take the form of an additional homework task rather than generate a debate on the consequences of the previous classroom activity.

### *Training in context*

A second factor concerns the context of the group work training. Some teachers admitted that they had skipped parts of the training, again on the grounds of time, and justified this action by arguing that similar training procedures were also part of the personal, social and health education (PSHE) programme. This raises questions concerning transfer of training where there is strong evidence that if transfer is to be effective it has to be contextualised (Good & Brophy, 2002) so that developing the rules for working in groups in PSHE may not the same as developing rules for working in groups in science. Indeed, there was evidence for this within the study in that in some schools the same pupils were observed participating in groups in both science and English where their levels of cooperation were noticeably different.

***Reporting back***

A third factor which tended to diminish the effectiveness of the groups involved the quality of feedback which often emerged during the 'reporting back' session. Built into most teachers' antennae is the notion of equity, particularly with respect to the opportunity to learn. Thus in the earlier ORACLE study of primary children and their transfer to the first year of secondary school, when at that time the most common form of teacher pupil interaction was through individual attention, it was noticeable that the quantity of teacher-pupil interaction received by each pupil was very similar (Galton *et al.*, 1980). This came about because even when teachers devoted an extra amount of time to a particular pupil on any one day they tended to compensate by giving more attention to other pupils subsequently. In the same way teachers in the present study appeared to feel that if children had produced work in groups then all groups should be given the same opportunity to tell the rest of the class what they had accomplished. Given the limited available time, these reporting-back sessions by the various groups tended to be short and to consist mainly of reporting what had been done and the results of this activity, rather than giving explanations for the group's chosen approach since this would have taken considerably longer. Yet the key purpose of using group work to develop the quality of pupils' thinking rests in the increased use of explanations as part of the reasoning process (Webb, 1988). Part of the value of group work is therefore lost if this kind of reasoning is not present and not shared by all the pupils during the reporting-back and debriefing stages.

Only in a few cases did teachers attempt to solve the shortage of time problem. In one particular case in a school well endowed with IT equipment, laptops were brought into the classroom and each group presented their experimental results and their reasoning in the form of a short Powerpoint presentation which was then emailed through a wireless connection to other groups. These presentations were then the object of the class discussion, with the result that a degree of repetition which often occurs when groups present their experimental findings was eliminated.

***Cognitive scaffolding***

The fourth factor which inhibited greater progress within the groups concerned the manner of scaffolding the various activities. There is good evidence to suggest that in teaching for understanding and for problem solving requiring metacognitive awareness, scaffolding should be built into the task itself rather than teacher directed (Rosenshine, Meister, & Chapman, 1996). Within the present teaching culture, however, teacher-directed scaffolding such as demonstration and guided discovery are promoted as the most effective means of support rather than the building supporting cues within the task, an approach often advocated in the cognitive developmental literature. Guided discovery, in particular, has come to be associated with pupil dependency (Doyle, 1986; Galton, 1989) in that pupils become adept at pushing teachers to give more and more guidance until they provide the required answers. Thus using Doyle's terminology the guided discovery approach reduces the risk of pupils failing by lowering the ambiguity of the task. When cues are incorporated into the task instructions, on the other hand, it becomes possible to maintain the task's ambiguity while reducing the risk of failure by framing tasks in such a way that pupils feel that their initial efforts have some relationship to what is

ultimately required. During observation of the group lessons it was noticeable that teachers gave pupils very little thinking time before they intervened within the groups and during subsequent interviews pupils often complained about this saying that they wanted to be left 'to work things out for themselves' until they had arrived as a group at some initial consensus. Then, when the teacher came to the group, they felt collectively strong enough to argue for their view. On the other hand, when teachers intervened too early, to give some guidance' pupils tended to see this as a sign of 'teacher take-over' in which he or she were seeking to impose certain of their own ideas. This was strongly resented.

### *The importance of 'neutral space'*

As a response to the problem of 'teacher take-over' some teachers in the study in a particular school created the idea of responding to pupils in what was termed 'neutral space'. As Neil (1997) has demonstrated both space and posture have major implications for the way that pupils interpret the teacher's actions within the classroom. While textbooks often show teachers operating at the same level as the pupils within the group (by kneeling at a table or sitting down on a chair) this, if done too early, may reinforce the notion of take-over because the adult is seen to be imposing himself within the group's space. On the other hand, the area near the teacher's desk or at the front of the class is one where it is customary for teachers to offer direct instruction and this space is therefore clearly seen by pupils as belonging to the teacher. Neutral space on the other hand (near the door, by the window, at the back of the class) can be conceived as belonging to neither teacher nor pupil and therefore does not carry the same overtones of ownership and direction. In the present study some English teachers devised a strategy where when joining a group to listen to the discussion they would often sit at an angle to the table so that they were not directly in eye contact with the group members. In this situation one teacher sometimes sat with her hand over her mouth, as if indicating that she was intending to listen but not speak. At some point when these teachers had visited most of the groups they would then go to a position of 'neutral' space, tell the class that that they had been listening to various ideas, that some of these were very excellent and that they would like to share them with all groups and also include one or two comments of their own. Subsequently when visiting the groups it was then possible to ask whether any of those ideas had been useful to the group and in this way it appeared that pupils were more ready to accept outside advice when attempting to reach a consensus view.

## Concluding comment: social pedagogy

Finally, as Kutnick *et al.* (2005b) have observed, theories of group work tend to divide into those which emphasise the social applications of collaborative working as opposed to those that emphasise academic outcomes. Because social relational elements tend to dominate, teachers are often not precise enough in defining their learning goals and in deciding whether these goals are worth the time and effort involved in setting up and managing group work effectively. All too often in this study because the purposes of working in groups were not clearly delineated, teachers failed to carry out the necessary activities such as briefing and debriefing to ensure the effectiveness of the procedures adopted. Thus if group work is to play a

greater part in the pedagogic repertoire of teachers there is need to put more emphasis on training, perhaps not so much at the initial teacher education level, where trainee teachers have enough to contend with, but rather in later professional development. Nevertheless, despite the limitations in current classroom practice described in the preceding paragraphs, the findings presented here are sufficiently encouraging to suggest that greater efforts with respect to professional training would pay rich dividends.

## Notes

1. Listed on the TLRP website as 'Improving the Effectiveness of Pupil Groups in Classrooms'.
2. Note: The effect sizes calculated in Tables 3, 4 and 5 are for correlated data and are computed according to Cohen (1988, pp. 48–49). Generally, the larger the correlation between two sets of scores the larger the effect size.

## Notes on contributors

Maurice Galton is Emeritus Professor of the University of Leicester and Senior Research Fellow in the Centre for Commonwealth Education at the University of Cambridge. He has directed two major studies of group work. The first in the 1990s took place in primary schools and the second, recently completed research involved pupils aged 11 to 14 years.

Linda Hargreaves is Reader in Classroom Learning and Pedagogy in the Faculty of Education at the University of Cambridge. She has collaborated with Maurice Galton on many classroom studies including the recent group work project with 11–14 pupils. She is a recognized expert in use of systematic observation to study classroom processes.

Tony Pell is a researcher in the School of Education at the University of Leicester. He is an authority on the construction of attitude inventories with a special interest in pupils' dispositions towards science. He has undertaken the analysis of most of Maurice Galton's projects including the latest on group work with 11–14 pupils.

## References

Alexander, R.J. (2006). *Towards dialogic teaching: Rethinking classroom talk*. York: Dialogos.

Brehony, K. (2005). Primary schooling under New Labour. *Oxford Review of Education*, *31*(1), 29–46.

Brinton, B., Fujiki, M., & Higbee, J. (1998). Participation in cooperative learning activities by children with specific language impairment. *Journal of Speech, Language and Hearing Research*, *41*, 1193–1206.

Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*(1), 1–35.

Cohen, J. (1988). *Power analysis for the behavioral sciences*. New Jersey: Lawrence Applebaum Associates.

Cowie, H., Smith, P., Boulton, M., & Laver, R. (1994). *Cooperation in the multi-ethnic classroom*. London: David Fulton.

Cowie, H., & Rudduck, J. (1988). *School and classroom studies*. Volume 2 of Learning Together-Working Together Project. London: BP (British Petroleum) Educational Service.

Croll, P. (1986). *Systematic classroom observation*. Lewes: Falmer Press.
DfES. (2003). *The Key Stage 3 Strategy, introducing the third year*. London: DfES.
DfES. (2004). *Pedagogy and practice: Teaching and learning in secondary schools*. London: DfES.
DfES. (2007). *Gender and education: The evidence on pupils in England*. London: DfES.
Dillon, J. (1990). *The practice of questioning*. London: Routledge.
Doyle, W. (1983). Academic work. *Review of Educational Research*, *53*, 159–199.
Dunkerton, J. (1981). Should classroom observation be quantitative? *Educational Research*, *23*, 144–151.
Edwards, D., & Mercer, N. (1987). *Common knowledge and the development of understanding in the classroom*. London: Routledge.
Eggleston, J., Galton, M., & Jones, M. (1975). *A science teaching observation schedule*. London: Macmillan and The Schools' Council.
Entwistle, N.J. (1977). Strategies of learning and studying: Recent research findings. *British Journal of Educational Studies*, *25*, 225–238.
Farivar, S., & Webb, N. (1991). *Helping behavior activities handbook: Cooperative group problem solving in middle school mathematics*. Los Angeles: UCLA.
Gage, N. (1978). *The scientific basis for the art of teaching*. New York: Teachers College Press.
Galton, M. (1989). *Teaching in the primary school*. London: David Fulton Publishers.
Galton, M., Hargreaves, L., Comber, C., & Wall, D. with Pell, T. (1999). *Inside the primary classroom: 20 years on*. London: Routledge.
Galton, M., & MacBeath, J. (2008). *Teachers under pressure?* London: Sage.
Galton, M., Simon, B., & Croll, P. (1980). *Inside the primary classroom*. London: Routledge.
Galton, M., & Williamson, J. (1992). *Group work in the primary classroom*. London: Routledge.
Good, T., & Brophy, J. (2002). *Looking in classroom*. Boston: Allyn & Bacon.
Hargreaves, L., Moyles, J., Merry, R., Patterson, F., Pell, A., & Esarte-Sarries, V. (2003). How do primary school teachers define and implement interactive teaching in the national literacy strategy in England? *Research Papers in Education*, *18*(3), 217–236.
Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, *43*(6), 387–425.
Kagan, S. (1988). *Cooperative learning: Resources for teachers*. Riverside, CA: University of California Press.
Kingsley-Mills, C., McNamara, S., & Woodward, L. (1992). *Out from behind the desk: A practical guide to groupwork skills and processes*. Glenfield: Leicestershire County Council Publications.
Kulik, J., & Kulik, C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, *36*, 73–77.
Kutnick, P., Blatchford, P., Clark, H., MacIntyre, H., & Baines, E. (2005a). Teachers' understandings of the relationship between within-class (pupil) grouping and learning in secondary schools. *Educational Research*, *47*(1), 1–24.
Kutnick, P., Hodgkinson, S., Sebba, J., Humphreys, S., Galton, M., Steward, S., Blatchford, P., & Bains, E. (2007). *Pupil grouping strategies and practices at Key Stage 2 and 3*. Research Report 796. Nottingham: DfES Publications.
Kutnick, P., Sebba, J., Blatchford, P., Galton, M., Thorp, J. with MacIntyre, H., & Berdondini, L. (2005b). *The effects of pupil grouping: Literature review*. Research Report 688. Nottingham: DfES Publications.
Lou, Y., Abrami, P., Spence, J., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within class groupings: A meta analysis. *Review of Educational Research*, *66*(4), 428–458.
MacBeath, J., & Galton, M. (2004). *A life in secondary teaching: Finding time for learning*. London: NUT.

Mercer, N. (2000). *Words and minds: How we use language to think together*. London: Routledge.

Neill, S. (1991). *Classroom non-verbal communication*. Abingdon: Routledge.

Ofsted. (2005). *Managing pupil behaviour*. London: Office for Standards in Education.

Pell, T., Galton, M., Steward, S., Page, C., & Hargreaves, L. (2007). Group work at Key Stage 3: Solving an attitudinal crisis among young adolescents? *Research Papers in Education*, *22*(3), 309–332.

Putnam, J., Markovchick, K., Johnson, D., & Johnson, R. (1996). Cooperative learning and peer acceptance of pupils with learning disabilities. *Journal of Social Psychology*, *136*, 741–752.

Rosenshine, B. (1979). Content, time and direct instruction. In P. Peterson & H. Walberg (Eds.), *Research on teaching concepts, findings and implications*. Berkeley, CA: McCutchan.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of intervention studies. *Review of Educational Research*, *66*(2), 181–221.

Sharan, Y., & Sharan, S. (1992). *Expanding cooperative learning through group investigation*. New York: Teachers College Press.

Slavin, R. (1983). When does cooperative learning increase student motivation? *Psychological Bulletin*, *94*(3), 429–445.

Slavin, R. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*, *57*, 293–336.

Stobart, G., & Stoll, L. (2005). The Key Stage 3 Strategy: What kind of reform is this? *Cambridge Journal of Education*, *35*(2), 225–238.

Stoll, L., Stobart, G., Martin, S., Freeman, S., Freedman, E., Sammons, P., & Smees, R. (2003). *Preparing for change: Evaluation of the implementation of the Key Stage 3 Pilot Strategy*. London: Department for Education and Skills.

TLRP (Teaching and Learning Research Programme). (2008). *Improving the effectiveness of pupil groups in classrooms 2001–4*. Retrieved November 13, 2008, from http://www.tlrp.org/proj/phase11/phase2a.html

Tomlinson, S. (2005). *Education in a post-welfare society*. Maidenhead: Open University.

Warrington, M., & Younger, M. (2006). *Raising boys' achievement: Towards an holistic approach*. Maidenhead: Open University Press.

Webb, N. (1985). Student interaction and learning in small groups: A research summary. In R. Slavin, S. Sharan, S. Kagan, S. Hertz-Lazarowitz, N. Webb & R. Schmuck (Eds.), *Learning to cooperate, cooperating to learn* (pp. 147–172) (New York: Plenum Press).

Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*(1), 21–39.

Webb, N. (1991). Task related verbal instruction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, *22*(3), 366–389.

Webb, N., Baxter, G., & Thompson, L. (1997). Teachers' grouping practices in fifth grade science classrooms. *The Elementary School Journal*, *98*(2), 91–113.

Webb, N., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal*, *31*, 369–395.

Webb, N., & Mastergeorge, A. (2003). Promoting effective helping behaviour in peer directed groups. *International Journal of Educational Research*, *39*(1/2), 79–97.

Wilkinson, J., & Canter, S. (1982). *Social skills training manual: Assessment, programme, design and management of training*. Chichester: John Wiley & Sons.

Younger, M., Warrington, M., & McLellan, R. (2005). *Raising boys' achievements in secondary schools: Issues, dilemmas and opportunities*. Buckingham: Open University Press.

**Appendix 1. NFER (National Foundation of Educational Research) conversion from national assessment task levels to point scores**

*A typical pupil aged 11 in Year 6, at the end of primary school, should reach Level 4, and a typical 14-year-old pupil in Year 9, at the end of lower secondary school, Level 5. A year's progress therefore equals 2 NFER points.*

| Typical pupil: age (year group) | Assessment task level and equivalent NfER points | Typical pupil: age (year group) | Assessment task level and equivalent NfER points |
|---|---|---|---|
| Age 8 (Year 3) | Level 3=21 | Age 14 (Year 9) | Level 5=33 |
| | Level 3+=23 | | Level 5+=35 |
| | Level 4−=25 | | Level 6−=37 |
| Age 11 (Year 6) | Level 4=27 | | Level 6=39 |
| | Level 4+=29 | | Level 6+=41 |
| | Level 5−=31 | | |