

Running head: Social effects of collaboration

Social effects of collaborative learning in primary schools

Andrew Kenneth Tolmie ^{a*}, Keith J. Topping ^b, Donald Christie ^c, Caroline Donaldson ^d, Christine Howe ^e, Emma Jessiman ^f, Kay Livingston ^g, Allen Thurston ^h

^a *Department of Psychology and Human Development, Institute of Education, University of London, 25 Woburn Square, London WC1H 0AA, United Kingdom*

^b *School of Education, Social Work and Community Education, University of Dundee, Nethergate, Dundee DD1 4HN, United Kingdom*

^c *Department of Childhood and Primary Studies, University of Strathclyde, Jordanhill Campus, Glasgow G13 1PP, United Kingdom*

^d *Centre for the Older Person's Agenda, Faculty of Health and Social Sciences, Queen Margaret University Edinburgh, Musselburgh EH21 6UU, United Kingdom*

^e *Faculty of Education, University of Cambridge, 184 Hills Road, Cambridge CB2 8PQ, United Kingdom*

^f *Fife Council Psychological Service, Auchterderran Centre, 14 Woodend Road, Auchterderran, Lochgelly KY5 0NE, United Kingdom*

^g *International Educational Consultancy, Faculty of Education, University of Glasgow, 11 Eldon Street, Glasgow G3 6NH, United Kingdom*

^h *Stirling Institute of Education, University of Stirling, Stirling FK9 4LA, United Kingdom*

Abstract

There is conflicting evidence on whether collaborative group work leads to improved classroom relations, and if so how. A before and after design was used to measure the impact on work and play relations of a collaborative learning programme involving 575 students 9-12 years old in single- and mixed-age classes across urban and rural schools. Data were also collected on student interactions and teacher ratings of their group-work skills. Analysis of variance revealed significant gains for both types of relation. Multilevel modelling indicated that better work relations were the product of improving group skills, which offset tensions produced by transactive dialogue, and this effect fed through in turn to play relations. Although before intervention rural children were familiar with each other neither this nor age mix affected outcomes. The results suggest the social benefits of collaborative learning are a separate outcome of group work, rather than being either a pre-condition for, or a direct consequence of successful activity, but that initial training in group skills may serve to enhance these benefits.

Keywords: Collaborative learning; Group work; Classroom relations; Work relations; Transactive dialogue

* Corresponding author. Tel.: +44 20 76126888; fax +44 20 76126304
E-mail address: a.tolmie@ioe.ac.uk (A. Tolmie)

1. Introduction

1.1. Cognitive and social gains in cooperative and collaborative learning

Group work in schools is generally regarded as falling into one of three types, peer tutoring, cooperative learning and collaborative learning—these being distinguished by increasing levels of equality and transactive engagement between students (Damon & Phelps, 1989; Foot, Morgan, & Shute, 1990). Thus cooperative and collaborative learning diverge from peer tutoring in their shared emphasis on horizontal rather than vertical variation in the knowledge and expertise held by the students comprising a group; and on discussion and exchange of information and ideas between students as a key mechanism promoting growth (Dillenbourg, 1999; Howe & Tolmie, 1998; Johnson & Johnson, 1979). They differ in turn from each other in the degree of transactive exchange they foster. Cooperative learning typically involves highly structured, wide-ranging programmes of activity, and makes use of jigsaw methods, in which students carry out individual tasks, and then share outcomes with other group members (Aronson, Stephan, Sikes, Blaney, & Snapp, 1978; Cohen, 1994; Sharan, 1980). Collaborative learning uses self-contained tasks and focuses solely on joint activity (Foot et al., 1990), typically with the overt objective of creating shared understanding, particularly where older learners are involved (Roschelle & Teasley, 1995; Schwartz, 1998; Summers, 2006).

This variation in emphasis stems from differences in theoretical background. Cooperative learning research originated in social psychological work on teams (Bales, 1950; Deutsch, 1949; May & Doob, 1937), and retains a focus on conditions that promote coordination of effort and understanding (Cohen, 1994; Slavin, 1989; Summers & Svinicki, 2007; Webb & Farivar, 1999). Collaborative learning, especially as applied to younger students, is founded in work on sociocognitive conflict (Doise & Mugny, 1984; Piaget, 1932, 1985), and research here has concentrated on how differences in viewpoint serve as a catalyst and a resource for conceptual change (Azmitia, 2000; Clark, Anderson, Kuo, Kim, Archodidou, & Nguyen-Jahiel, 2003; Howe & Tolmie, 1998).

Despite these differences, cooperative and collaborative learning are the subject of similar claims regarding the benefits that they are capable of engendering across the primary and secondary school age range, and on into university education. These include improvements in participants' conceptual grasp and application of skills (Gillies & Ashman, 2003; Howe & Tolmie, 1998; Johnson & Johnson, 1979), but also more positive social relations (Azmitia, 2000; Blatchford, Baines, Rubie-Davies, Bassett, & Chowne, 2006; Marks, 2000). This dual impact on achievement and social integration is of considerable practical significance, since it makes it easier to justify implementation of group work in crowded curricula.

1.2. Possible relationships between cognitive and social gains

Despite these claims, it is apparent on closer analysis that the relationship between cognitive and social gains from group work is not straightforward. In general, there is better evidence for a dual impact in the context of cooperative learning, where implementation is usually more extensive and resource-intensive. An early meta-analysis by Johnson, Maruyama, Johnson, Nelson, and Skon (1981), examining the effects of cooperative, competitive and individualistic teaching methods on student achievement, revealed that cooperation consistently led to higher

achievement and motivation to learn across a range of subjects than either competition or individual study. A subsequent meta-analysis by Johnson, Johnson, and Maruyama (1983) found that cooperative learning was also associated with more positive perceptions of other class members, including those from different ethnic groups and individuals with disabilities. A more recent meta-analysis of 148 studies from 11 countries and a wide range of social backgrounds (Roseth, Fang, Johnson, & Johnson, 2006) confirmed both findings for middle school students aged 12 to 15 years, and reported that achievement was strongly related to interpersonal perception.

However, these analyses provide no clear evidence as to whether good relations between group members are: (a) a pre-condition for achievement, and thus a crucial first step in any effective cooperative learning programme; (b) a directly-related consequence of the processes giving rise to positive achievement; or (c) a separate outcome of some element of those processes, and therefore possibly not guaranteed to occur in all instances. Taking social interdependence theory (Deutsch, 1962) as a basis, Roseth et al. (2006) argue for a version of (b). They hold that cooperative learning leads to a form of positive feedback loop, in which perceptions of group members are enhanced by successful activity, and this further facilitates subsequent cooperation. Similarly, Summers and Svinicki (2007) report that amongst university students exposed to cooperative learning, perceptions of classroom community were related to how successful they felt their groups were at working collectively. Nevertheless, the alternative relationships specified by (a) and (c) cannot be ruled out in the absence of an analysis of the pattern of change in both dimensions that occurs as a result of cooperative learning.

Uncertainties about the nature of the dual impact in the context of cooperative learning become even more pronounced with respect to collaborative learning, an important point, given that it is an approach that is often preferred by teachers as something that can be readily used alongside whole-class activities (Blatchford et al., 2006; Galton, Hargreaves, Comber, Wall, & Pell, 1999). There is little evidence of dual effects here, and the emphasis on progress through conflict may mean that positive social relationships are a pre-condition which needs to be established for achievement gains to occur, rather than a related outcome. There is certainly substantial evidence that the effects of collaborative learning on achievement stem primarily from students proposing and explaining ideas to each other (Howe, Tolmie, Anderson, & Mackenzie, 1992; Tolmie & Howe, 1993; Tolmie, Howe, Mackenzie, & Greer, 1993), and that discussion of this kind is commonly precipitated by differences in perspective and explicit disagreement (Doise & Mugny, 1984; Howe & Tolmie, 1998; Mercer, 1995). Moreover, in line with Piaget's account of change through internal disequilibrium and re-equilibration, progress does not seem to depend on within-group resolution of these disagreements – which is indeed commonly absent, amongst younger learners at least – although it does not hamper it (Howe, Tolmie, & Rodgers, 1992; Piaget, 1985; Williams & Tolmie, 2000). This could be problematic both socially and intellectually, in that unresolved differences might fuel resentment and curtail further interaction unless participants were tolerant of their occurrence from the outset.

Support for the pre-condition argument is provided by evidence from Azmitia and Montgomery (1993) and MacDonald and Miell (2000). Both studies compared the performance of pairs of 11-year-old friends to that of acquaintances, in the first case on an isolation of variables task in elementary science, and in the second on a music composition task. Friendship pairs performed better on both tasks, exhibiting more transactive dialogue, and a greater readiness to critique and elaborate on

partners' ideas, but these advantages were apparent only when the task was more challenging. MacDonald and Miell (2000) argue this is because friends feel more confident about revealing ideas and criticising each other, especially under conditions of greater uncertainty. This line of reasoning is strongly reminiscent of Edmondson's (1999) construct of psychological safety, defined as a perception on the part of an individual that they will suffer no negative consequences for their self-image or status if they express their thoughts and feelings. Whilst originally devised to account for workplace behaviour, Van den Bossche, Van Gennip, Gijssels, and Segers (2005) argue that psychological safety is a prerequisite for constructive engagement in collaborative learning, and report that ratings of safety amongst collaborating university students were associated positively with constructive conflict, and negatively with other disputes. Arvaja, Hakkinen, Rasku-Puttonen, and Etelapelto (2002) similarly report that an increased fluidity of collaboration is apparent where high school students' relationships are characterised by high levels of mutual trust.

The evidence that positive social relations facilitate productive exchange seems compelling. However, that their absence results in social friction – a key requirement of the pre-condition account – is less clear. Issroff (1995) found that for some 13- to 14-year-olds collaborative work was sufficiently negative an experience to leave them with substantial hostility towards their partner, even though the outcome was positive in terms of learning. Roschelle and Teasley (1995) report that when pairs of 15-year-olds collaborating on a physics task encountered a breakdown in mutual intelligibility, this led to disengagement from both the task and dialogue with partners, if only temporarily. Shachar and Fischer (2004), examining the impact of an extended programme of collaborative work, found that whilst this had positive effects on learning, it led to a decline in motivation to participate in such work, perhaps reflecting negative effects on class relations. Since no attempt was made in any of this research to establish positive social relations at the outset, any resulting friction would be entirely understandable from the perspective of the pre-condition account. There is little evidence of negative effects beyond these studies, though, and on current data, such outcomes seem to be neither widespread nor consistent.

It is possible then that good pre-existing social relations serve to help optimise collaboration rather than to determine whether it is workable. If so, under more general circumstances, the experience of collaborative activity over time might provide a means for students to build up positive relations alongside achievement gains, in line with the directly-related consequence or separate outcome accounts. Indeed, it is hard to see how else positive non-friend working relationships (i.e., those not based on other aspects of attraction) could emerge ordinarily. Support for this position is provided by Azmitia (2000), who argues that negative outcomes of collaborative work for social relations are relatively unusual, because children are socialised into wider conventions regarding expression and management of negative affect, and learn through experience to use these deliberately and productively. One such convention is to take a 'time-out' or to disengage from the interaction for a space (cf. Roschelle & Teasley, 1995), allowing negative feelings to dissipate before re-engaging. Kreijns and Kirschner (2005) also frame the social consequences of collaborative work in positive terms. They argue that group learning has a socio-emotional as well as socio-cognitive dimension, directed at the formation of what Rourke (2000) terms a 'sound social space'. Both arguments imply that experience of collaborative engagement helps learners to acquire relational skills with an impact beyond the immediate group context; and that this is a separate outcome to any cognitive gain.

One caveat is indicated by the work of Buchs and colleagues (Buchs, Butera, Mugny, & Darnon, 2004; Darnon, Buchs, & Butera, 2002), who report differences in the impact of ‘epistemic’ and ‘relational’ solutions to conflicts during collaboration. Epistemic solutions, which focus on the task, led to better learning and more positive perceived relationships, whereas relational solutions, which focus on the relative competence of participants, were associated with worse learning and more negative relationships than controls. Jarvelä, Lehtinen, and Salonen (2000) identify individual differences in emotional orientation to tasks, especially at points of challenge, which might underlie the tendency to seek one or other type of solution to group conflicts. The implication is that whilst positive relationships are not a pre-condition, social growth from collaborative learning is more likely if this activity is preceded by a period of preparation, during which explicit ground rules for dealing with conflicts in epistemic fashion are established.

In line with this, Wegerif (2000) argues that successful collaboration amongst school children requires preparation for the management of discussion, including acceptance of disagreement (see also Dillenbourg, 1999). Direction of initial group exchanges by researchers (see Howe et al., 1992) may have served this purpose in past experimental work, perhaps explaining the limited reports of collaboration having negative effects. Kutnick and Manson (1998) argue that broader relational preparation is needed too, to provide a context within which productive discursive conventions are more likely to take root. Kutnick, Ota, and Berdondini (2008), and Blatchford et al. (2006) both demonstrate the benefits of such preparation on elementary school students’ engagement in group work and levels of communicative interaction.

1.3. Aims of the present research

Taken overall, then, rather than positive relationships being a pre-condition for cognitive progress, there are grounds for thinking that collaborative group work can benefit perceived relationships with classmates at the same time as improving conceptual grasp, provided students are given some initial relational and communicative preparation. The precise manner in which the processes leading to cognitive and social gains articulate with each other remains less clear, however. These may be separate outcomes of collaborative discussion (cf. Kreijns & Kirschner, 2005), with successful navigation through transactive exchanges promoting social gains, whilst the content of these promotes conceptual change. Alternatively, they may be directly-related consequences of the process of exchange, as the data on the impact of epistemic conflict resolution (Darnon et al., 2002) would seem to imply (cf. also Roseth et al., 2006, on the social interdependence effects generated by successful achievement).

It is also unclear whether contextual differences might act to moderate social gains, given that classrooms vary considerably in terms of background social relationships. For instance, compared to single-age classes, mixed-age classes may engender more distant relationships, characterised by deferral (cf. Bachmann & Grossen, 2004; Piaget, 1932), making increases in positive relations harder to achieve. On the other hand, children in classes drawn from the same small rural community are likely to have better-established non-work relationships than those in urban classes (Bell & Sigsworth, 1987). They may therefore find it easier to develop more positive work relationships even if non-work relations show less benefit, being stronger already; see Shapira and Madsen (1969) on greater cooperative tendencies within school of children from more collective communities.

The analyses reported in the present paper were therefore intended to (a) examine the general prediction that initially-supported collaborative group work leads to social gains alongside achievement gains; (b) investigate the precise relationship between social and cognitive gains; and (c) establish whether variation in social context affects outcomes. In order to do this, data from a group-work intervention that led to confirmed achievement gains (Howe, Tolmie, Thurston, Topping, Christie, Livingston, Jessiman, & Donaldson, 2007) were examined for evidence of change in classroom relationships, and for effects on these of social context and group activity. This intervention involved teachers and students from single- and mixed-age upper primary (elementary) classes, drawn from separate schools in rural and urban areas, who participated in a six-month collaborative group-work programme. This consisted of an initial phase of generic relational and communication skills training, followed by structured collaborative work (i.e., a series of self-contained joint activities) in two science topics, all classes utilising the same pre-prepared materials in both phases. To promote implementation quality, teachers received training immediately prior to each phase, but beyond this, use of the materials with which they were provided was left free to vary. As is reported in detail by Howe et al. (2007), analysis of the intervention's cognitive impact showed that the participating students exhibited significant gains in understanding of the two science topics. In both instances, these gains were a direct function of the scale of transactive exchange (proposition and explanation of ideas) that took place during group activity, and of the extent of teacher support for such dialogue.

To investigate social gains, the research utilised a simple before and after design, along with an analysis of the processes leading to observed change. This design was necessitated by difficulties in recruiting control classes (see Howe et al., 2007), especially with respect to the provision of relationship data. Measurements of work and social relations among classmates were taken before and after intervention (bearing in mind the predicted differences in impact of collaborative activity on these among rural students). These measures focused on the number of other class members held in positive regard, so that the impact of collaborative work on broad classroom 'temperature' could be assessed (cf. Kreijns & Kirschner, 2005; Roseth et al., 2006). In order to establish the quality of the group work that took place, and whether it improved over the course of the intervention (cf. Kreijns & Kirschner on socio-emotional learning), observations of on-task dialogue were taken at intervals, along with measures of the more general character of group activity within each class. These data, along with assessments of group-work skills and broader impact provided by teachers, made it possible to examine influences from group interaction on pre- to post-intervention change in social relations.

1.3.1. Hypotheses

It was hypothesised that, contrary to the pre-condition account, students would exhibit gains pre- to post-intervention in the number of classmates with whom they liked working and playing (Hypothesis 1), but that these gains would be subject to the effects of social context. In particular, it was anticipated that gains would be greater in single-age classes, and also in rural classes as far as work relationships were concerned (Hypothesis 2). With regard to the process of change, it was argued that if social gains were found to be positively associated with transactive dialogue (that is, the main predictor of cognitive gains) then this would provide support for the hypothesis that social benefits are a directly-related consequence of the process that produces achievement (Hypothesis 3). If social gains were related primarily to indices

of the management of group work and not to transactive dialogue itself, then this would provide support for the hypothesis that cognitive and social benefits are separate outcomes (Hypothesis 4).

2. Method

2.1. Sampling

Participating schools were recruited via a survey on use of group work that was issued to a randomly selected one-in-three sample of primary (elementary) schools in eight out of the 32 local authorities (school districts) in Scotland. Four of the target authorities were in the east of Scotland, and four in the west. Altogether 221 surveys were issued and 85 (38%) returned. From those schools expressing interest in taking further part in the study, a final sample of 24 classes in the last two years of primary education (ages 10-12 years) was selected, each from a separate school, balancing geographical and demographic variables. The characteristics of this sample are shown in Table 1.

 INSERT TABLE 1 ABOUT HERE

The classes in each section of the sample were divided evenly between schools in the eastern and western authorities. The distinction between rural and urban schools was made using the criterion employed by government in Scotland: whether or not the local population was above or below 10,000. The rural schools had significantly lower rolls than the urban ones ($M = 124.2$, $SD = 89.8$, and $M = 326.9$, $SD = 118.9$, respectively), $F(1, 20) = 23.97$, $p < .001$, but they were otherwise broadly comparable. In particular, all were located in socially mixed areas, with an overall 19.75% of attending children receiving free school meals (a standard index of relative deprivation in the UK), 21.42% in urban schools, 15.34% in rural ($p = ns$); and only a small number (1.8%) of children in the sample were from ethnic minority backgrounds. Total N was 575, but absences at pre- or post-testing meant that subsequent analyses commonly featured fewer children.

2.2. Nature of the intervention

2.2.1. Phase 1: Group-work skills training

Teachers of the 24 classes involved in the intervention attended three days of professional development training, and in addition made occasional use of email contact both with each other and with project staff to discuss issues that arose during the course of the intervention, primarily with regard to group composition and use of the science materials. Researcher visits to schools provided a further opportunity for informal support on these points. The first training day focused on the potential benefits of group work, principles for facilitating effective peer collaboration, and classroom activities designed to enhance children's generic relational and communication skills. A comprehensive package of support materials (see Blatchford, Galton, Kutnick, & Baines, 2005; Baines, Blatchford, Kutnick, Chowne, Ota, & Berdondini, 2008), including details on principles and strategies for achieving these, was issued to each teacher. These materials emphasised the importance of four main elements: classroom organisation, including appropriate seating arrangements; the use of tasks that encouraged collaboration by promoting discussion around joint activity;

the role of the teacher as a “guide on the side”, encouraging and modelling effective communication between students; and the nature of the relational and communicative skills students required to engage in collaborative group work. They also included plans for a sequence of group activities designed to help students build up mutual trust and tolerance, agree rules for working together, and develop advanced communication skills such as listening, speaking, explaining and compromising (see Baines et al. for details of these activities). Teachers were encouraged to devote one hour per week to generic group-work training activities over the ensuing 12-week period, using the resources provided, and to allocate a further hour per week to collaborative work within ‘normal’ curricular contexts, using groups of four to five students. Since some “stretching” of behaviour by working with different peers was likely to be needed to yield social benefits (cf. Polvi & Telama, 2000), teachers were asked not to form groups based on existing friendships, but were otherwise left free to determine composition.

2.2.2. Phase 2: Group work in science training

The second training day some 12 weeks later focused on curricular applications of collaborative work in the context of two science topics, Forces and Evaporation. It included practical experience of collaborative group work among the teachers in science activities related to these topics (partly to ensure correct understanding of the underlying scientific principles). Subsequently, teachers were expected to devote at least an hour per week over a 6- to 8-week period to structured group work on the two topics, supported by further comprehensive classroom resources (Baines et al., 2008; Howe et al., 2007). These detailed a series of activities, each focused on a concrete task, typically involving testing and accounting for particular phenomena. These tasks were designed to promote proposition and explanation of ideas by group members, via instructions to share and discuss their thoughts, reach a consensus, and record an agreed explanation in writing (cf. Howe & Tolmie, 1998, 2003). In addition to these activities, teachers were again expected to try to extend their use of group work across the rest of the curriculum. They were advised to maintain stable group membership throughout, as far as possible, in order to allow students to develop relationships with each other, although it was accepted that some fluctuation was inevitable. A final training day after the science activities had been completed was designed to provide teachers with the opportunity to engage in structured reflection on their experiences, and the lessons they had learnt about the use of collaborative work, as well as permitting the collection of more formal data on their evaluations of the intervention programme.

2.3. Instrumentation

2.3.1. Class relationships

An instrument entitled “People in your Class” was developed to provide individual-level data on children’s relations with classmates in school, in terms of who they liked working with in class and playing with during school breaktimes. This allowed measurement of change in these key dimensions pre- to post-intervention. The pre-intervention version also included four questions on out-of-school relationships, in order to ascertain that children attending rural and urban schools differed in the depth and breadth of these in the expected manner. The instrument itself was a variation of the Roster and Rating Scale (Roitacher, 1974), which presented a matrix listing all members of the respondent’s class (rows) and two

contexts (columns for “Like to play with at school breaktime” and “Like to work with in class”; plus pre-intervention “Play with out of school”, “See at local clubs or out-of-school classes”, “Know family of”, and “See at local events”). Children were asked to work down the list of names, simply placing a tick in the columns for the categories that applied to each person. Since collaborative groupings (and most of these types of out-of-school contact) were unlikely to be gender-restricted, no effort was made to constrain responses to same-gender classmates. In order to correct for variation in class size, individuals’ responses for each context were scored in terms of the proportion of those in their class whose names they had ticked, yielding six variables in total pre-intervention, and two post-intervention. It is important to note that although the pre-intervention ratings for out-of-school contexts were designed to capture extent of acquaintance rather than liking, taken together the six contexts did form a homogenous scale (Cronbach’s $\alpha = .69$), indicating that liking was at least partially related to contact. At the same time, responses to the two liking contexts measured both pre- and post-intervention were relatively distinct from each other, with only a small overlap in variance (pre-intervention, $r = .23$, $n = 539$, $p < .001$, but $r^2 = .05$).

2.3.2. Group work: Dialogue

Observations of classroom dialogue were made using a time-sampling method focused on the incidence of 12 elements associated with productive outcomes in previous research (Howe & Tolmie, 1998; Tolmie, Thomson, Foot, Whelan, Morrison, & McLaren, 2005; for greater detail on these codes see Howe et al., 2007). These observations were recorded *in situ* by researchers at three time points: during generic group-work training, during group work on evaporation, and during group work on forces. At each time point, two observation sessions were undertaken, one during a lesson employing group work, and the other one during a conventional class lesson, to provide a baseline for comparison. Visits were arranged at short notice, on the basis of teachers’ reported timetable plans, to reduce any possibility of observed lessons receiving special attention in terms of teachers’ preparations.

Prior to the first visit, six children per class, three girls and three boys, were identified at random from the class list to serve as observation targets, these same children (the majority of whom turned out to be members of different collaborative groups) then being used for all six observation sessions. During these sessions, each target child was unobtrusively observed for eight consecutive 40-second time windows (12 seconds to focus in, 16 seconds to observe and 12 seconds to record), before attention turned to the next child on the observer’s list. Within lessons employing group work, children were only observed when they were supposed to be conducting activities involving group work (i.e., not during preparation or summing up). Observations were recorded via ticks on a grid, with rows corresponding to sampling period, and columns to codes. Where appropriate, multiple codes were used within the same time window, so that the final record consisted of a measure of the actual frequency with which dialogue elements of different types occurred within the same fixed period of time for each child and for each lesson type.

The researchers who conducted the observations were trained in use of these grids prior to the initial sessions by applying them to previously recorded videotapes of interaction in primary science classrooms, and discussing the resultant coding until both were confident in its use. Inter-rater reliability was then checked via independent coding of 64 forty-second extracts from further videotapes of the same character. Agreement over coding to different categories was 92%. Subsequent factor analysis

(see Howe et al., 2007) indicated that the incidence of two of the 12 dialogue elements, propositions (suggested ideas or actions) and explanations (reasons for propositions) within the group-work in science lessons (i.e., during the second and third observations of group work) formed a conjoint stable index, which was found to predict cognitive gains. The total frequency of these two codes across those observations was therefore taken as the central measure of transactive dialogue in subsequent analyses.

2.3.3. *Group work: Management of interaction*

Wider, class-level measures of the quality of group activity and its management by both teachers and students were taken at the conclusion of each observed group-work lesson (i.e., on three occasions for each class), using a multiple-item index devised by Blatchford et al. (2005) for the Social Pedagogic Research in Groupwork (SPRinG) project. This index, referred to as “SPRinG – Teaching Observation Protocol” (S-TOP), was used by researchers to capture on a 3-point rating scale (not true, partly true, very true) the presence during their observations of key features relating to (a) the quality of the learning context (S-TOP Learn: 4 items); (b) the suitability of tasks and activities (S-TOP Tasks: 7 items); (c) the nature of adult (teacher) involvement (S-TOP Adults: 9 items); and (d) the group-work skills displayed (S-TOP Group Skills: 11 items). Items included: “The size of groups maximized student-student interaction” (S-TOP Learn); “The group-work task warranted the use of exploratory talk/discussion (suggestions, explanations, conjecture etc.)” (S-TOP Tasks); “The teacher modelled good interaction skills” (S-TOP Adults); “Students showed good conversational skills (e.g., active listening, no interruption, presenting a line of argument, etc.)” (S-TOP Group Skills). Ratings for items in the different subscales were totalled separately at each time point, higher scores indicating greater evidence of the relevant features in a given class. Cronbach’s alpha for these subscales was .67, .81, .74, and .87, respectively. Cumulative measures of the quality of group work (and in this sense, of the implementation) were also derived from these subscales, by computing the total rating for each subscale across the three points of observation.

2.3.4. *Teachers’ perceptions: Impact on class*

As part of activity during the final training day, teachers completed an evaluation questionnaire, seven items of which explored their perception of the impact of the group-work programme on children in their class. These items, which covered positive and negative changes in students’ group skills and motivation for group activity (e.g., “The children have been able to transfer their collaborative group-work skills to other curriculum areas”), were rated from 1 (not at all agree) to 4 (completely agree). Ratings were reversed as necessary to give common polarity, and totalled to give a single, class-level score, with higher being better. Cronbach’s alpha across the seven items was .77.

2.3.5. *Teachers’ perceptions: Collaborative learning skills*

Towards the end of the intervention, teachers also completed a 17-item scale (Collaborative Learning Evaluation Form, CLEF; Topping, 2003) rating the progress in collaborative learning skills demonstrated by their class as a whole. Items offered three response alternatives, scored 0 to 2 (no progress made, some progress made, substantial progress made), and covered similar issues in part to the S-TOP Group Skills subscale. However, there was less focus on specific management of dialogue,

and more on the wider organisation of group activity (e.g., “Do the groups take turns and help everyone to join in?”). The two measures also differed in terms of emphasis, with S-TOP assessing absolute performance and CLEF assessing improvement. Cronbach’s alpha across the 17 items was .89 and no item was out of step, indicating high reliability. The measure used was the total across items.

2.4. Data-gathering procedure

The project took place over the greater part of a full school year (September to June). Participating schools were recruited in September/October. Pre-testing took place in late November/early December, shortly after the first teacher training session, and immediately before initiation of generic group-work training. The group work in science lessons took place in March/April, with post-testing following in May/early June, after the science activities had been completed. The inter-test period was thus approximately six months, including school vacations. The People in your Class scale measuring the children’s relations with classmates was administered to children by their class teachers, but scored by researchers. Teachers were provided with notes on the administration procedure which acknowledged the complexity of the measure, and, to avoid potential confusion, asked them to steer students towards completing it by working down the list of names, placing a tick in the columns for each category that applied to a given person. The classroom observations were conducted by researchers. The teacher evaluations of impact and improvement in collaborative skills were self-completed.

2.5. Analysis

Preliminary analysis focused on establishing that children attending rural and urban schools differed as expected in their degree of out-of-class acquaintance prior to the intervention. To this end, a one-way MANOVA (rural vs. urban area) was conducted, taking the four out-of-school indices for which data were collected at pre-test as dependent variables. MANOVA was chosen as a method of analysis here to control the familywise error rate, which would otherwise be inflated by the use of multiple tests. In order to ascertain whether implementation quality was a potential source of influence on outcomes, this was also examined at this stage, in terms of both variation among individual classes and effects of social context. A two-way class-level MANOVA (Area x Class Type) was carried out to test for the latter, using the four cumulative S-TOP observation measures, and the teacher ratings of impact and progress in collaborative learning skills as dependent variables. Effect sizes were calculated as partial eta-squared in the MANOVAs, and in subsequent ANOVAs, as a meaningful index of the proportion of explained variance.

Attention then turned to how far students exhibited (a) improvement across the observation sessions in levels of transactive dialogue and reported group-work skills; and (b) pre- to post-intervention gains in work and play relations with classmates. Three-way ANOVAs (Session or Pre/Post x Area x Class Type) were employed to examine progress, and to check for predicted effects of social context on relationship gains. Moderating effects of age and gender were also considered at this stage via correlations and unrelated t-tests. All ANOVA results for gains were based on cases with a complete record, that is, data relating to all observation sessions, or from both pre- and post-intervention measurements, as appropriate. Results of follow-up tests

were only noted where they were significant following Bonferroni adjustment for multiple tests.

Finally, the factors influencing adjusted change in class relations (i.e., change controlling for pre-test variation) were examined via the use of multilevel (hierarchical) linear modelling (MLM), in view of the nesting of students within classes, and the class-level nature of the broader observational and rating measures. These analyses were run using HLM 6 (Raudenbush, Bryk, & Congdon, 2005), and were restricted to the sub-sample for whom dialogue data were available ($N = 130$), so that the impact of the established predictors of cognitive gains (frequency of propositions and explanations during group work in science lessons, and S-TOP Adults; see Howe et al., 2007) could be examined alongside other indices of the quality of group work. It was only possible to compute two-level models (student and class) using these data, since student and group levels were effectively synonymous, the majority of cases being the sole representative of the group within which they had worked. Final models were determined by means of an unbiased goodness-of-fit statistic, the Akaike Information Criterion (AIC; Akaike, 1974), where smaller values indicate better fit, and comparison of parameter estimates where competing models had similar AIC values. Whilst no formal effect size statistics have been agreed as yet for use within MLM, approximate estimates of effect size were derived from calculations of the percentage reduction in final models of unexplained variance at class and student levels, relative to the basic intercept and slope model for that data (see Luke, 2004). Results are reported in the following section in the order outlined above.

3. Results

3.1. *Urban vs. rural differences in pre-intervention out-of-class relationships*

As can be seen in Table 2, rural and urban children exhibited a consistent, and significant, pattern of pre-intervention difference in out-of-class acquaintance, as anticipated, Pillai's trace = .26, $F(4, 534) = 47.87$, $p < .001$, partial $\eta^2 = .26$. Follow-up ANOVAs revealed that although within-group variability was high, rural children were significantly more likely to play with each other out of school, $F(1, 537) = 31.84$, $p < .001$, partial $\eta^2 = .06$, to see each other at local clubs or out-of-school classes, $F(1, 537) = 24.47$, $p < .001$, partial $\eta^2 = .04$, to know each other's families, $F(1, 537) = 86.33$, $p < .001$, partial $\eta^2 = .14$, and to see each other at local events, $F(1, 537) = 173.91$, $p < .001$, partial $\eta^2 = .24$. Whilst the differences in general familiarity indicated by knowing each others' families and seeing each other at local events were more pronounced than those in contacts more likely to involve developed relationships (i.e., playing with each other out of school, and seeing each other at local clubs or out-of-school classes), there was clear evidence that, compared to urban children, rural classmates had better-grounded relations with each other. It was possible therefore that these might impact positively on work and play relationships in school, with which they were intercorrelated, as noted above.

 INSERT TABLE 2 ABOUT HERE

3.2. *Variation in quality of implementation*

Data on the measures of implementation quality are shown in Table 3. The four cumulative S-TOP observation measures all showed a strong bias towards values in the upper part of the range, but with a negative skew attributable to a tail containing cases with lower values. Teachers' ratings of class impact and of progress in group-work skills (CLEF scale) exhibited a similar picture. However, MANOVA identified no systematic main or interaction effects on these variations involving rural vs. urban context or class type: for rural vs. urban context, Pillai's trace = .38, $F(6, 14) = .28$, ns; for class type, Pillai's trace = .79, $F(6, 14) = .59$, ns; for the interaction between these Pillai's trace = .73, $F(6, 14) = .63$, ns. The researcher ratings were intercorrelated, $r =$ from .53 to .67, $p < .01$ (for all, two-tailed), as were the teacher ratings, $r = .56$, $p < .01$ (two-tailed), indicating that cases with lower values tended to have these across the different indices. Correlations between the two sets of measures were positive but nonsignificant, perhaps unsurprisingly given that the researcher ratings were formative and absolute, whereas the teacher ratings were summative and focused on relative progress.

 INSERT TABLE 3 ABOUT HERE

Taken overall, then, it would appear the quality of implementation was generally good, with the majority of classes scoring highly from the outset in terms of the presence of key features (values close to maximum were only possible if high scores were obtained at all three time points), and most teachers concurring that the positive impact of the implementation was close to ceiling. There was nevertheless sufficient variability in the ratings to make it possible to consider whether differences in implementation affected social outcomes. The group skills measures in particular had lower means relative to the maximum, and values close to a normal distribution, indicating that teachers found it easier to put the basic resources of appropriate classroom organisation (S-TOP Learn), tasks (S-TOP Tasks) and support (S-TOP Adults) in place than to ensure high levels of student group-work skills.

3.3. Improvements in transactive dialogue and group-work skills

As noted above, the total frequency of propositions and explanations across the two observations of group work in science lessons was taken as the principal measure of transactive dialogue for the purposes of examining influences on social gains, in view of its established association with cognitive gains (Howe et al., 2007). However, in order to establish whether participating students showed improvements in levels of transactive dialogue during the course of the intervention, separate totals were also computed for each of the three observation points. These totals showed a large degree of individual variation, but also a roughly linear increase of 75% between observation of the generic group-work training and the second observation of the group work in science lessons (for Time 1, $\underline{M} = 2.39$, $\underline{SD} = 2.32$; for Time 2, $\underline{M} = 3.14$, $\underline{SD} = 2.71$; for Time 3, $\underline{M} = 4.18$, $\underline{SD} = 3.53$), $F(2, 250) = 27.74$, $p < .001$, partial $\eta^2 = .18$, $n = 129$ for the sub-sample of observed children for whom data from all three time-points were available). The class-level S-TOP ratings of students' group-work skills also showed a significant increase over the same period, amounting to 37% of the effective scale (minimum = 11, maximum = 33; for Time 1, $\underline{M} = 25.75$, $\underline{SD} = 4.37$; for Time 2, $\underline{M} = 28.00$, $\underline{SD} = 3.45$; for Time 3, $\underline{M} = 31.17$, $\underline{SD} = 2.99$; $F(2, 40) = 25.12$, $p < .001$, partial $\eta^2 = .56$, $n = 24$ classes).

Participants therefore showed sizeable gains in group-work skills and levels of transactive dialogue following initial group training. Neither gain was affected by classroom context, save that initial variations in group-work skills reduced as progress was made. At the first time point, skills tended to be rated as better in rural mixed-age ($M = 28.83$, $SD = 1.53$) and urban single-age classes ($M = 27.17$, $SD = 1.53$) compared to rural single-age ($M = 25.00$, $SD = 1.53$) and urban mixed-age classes ($M = 22.00$, $SD = 1.53$). By the third time point, however, the range in the values of the means had decreased substantially, from 6.83 to 2.83 (for rural mixed-age, $M = 32.50$, $SD = 1.19$; for urban single-age, $M = 30.33$, $SD = 1.19$; for rural single-age, $M = 32.17$, $SD = 1.19$; for urban mixed-age, $M = 29.67$, $SD = 1.19$). This convergence gave rise to a three-way interaction between time, area, and class type, $F(2, 40) = 3.78$, $p < .05$, partial $\eta^2 = .16$. Moreover, as class levels of group-work skill increased, they became positively related to the improved individual levels of transactive dialogue (at Time 3, $r = .17$, $n = 136$ children for whom observational data were available at this point, $p < .05$, one-tailed), suggesting better group-work skills helped facilitate productive exchange.

3.4. Pre- to post-intervention social gains

3.4.1. Work relations

Table 4a shows the average proportion of classmates students indicated they liked working with pre- and post-intervention, broken down by rural vs. urban and mixed- vs. single-age classes. Values are also given for pre- to post-intervention change. As can be seen, there was a general tendency for the number of classmates regarded as desirable work partners to increase pre- to post-intervention, though there were fluctuations in both the actual proportion and degree of change depending on area and class type. Overall pre- to post-intervention change was highly significant, though the effect size was not large, $F(1, 477) = 22.03$, $p < .001$, partial $\eta^2 = .04$, but there was also a significant interaction between change and class type, $F(1, 477) = 11.10$, $p = .001$, partial $\eta^2 = .02$. Children in single-age classes showed larger improvements in work relations than those in mixed-age classes ($M = .10$, $SD = .29$, $t(233) = 5.24$, $p < .001$, and $M = .01$, $SD = .25$, $t(246) = .92$, *ns*, respectively). However, mixed-age classes had better work relations pre-intervention, a difference which to some extent persisted post-intervention, resulting in an overall main effect of class type, $F(1, 477) = 15.59$, $p < .001$, partial $\eta^2 = .03$.

The greater gains exhibited by children in single-age classes were thus due primarily to them catching up with children in mixed-age classes, as if the intervention acted in some respect to provide single-age classes with an influence on work relations corresponding to one occurring more naturally in mixed-age classes. There was also a weak main effect of area, $F(1, 477) = 6.12$, $p < .05$, partial $\eta^2 = .01$, with rural children enjoying on average better work relations both pre- and post-intervention than urban children, consistent with the evidence of better-grounded initial relations reported above. Rural children showed no greater gains than urban, though, and in general the effects of class type outweighed those of broader social context. It should be noted finally that there was a substantial level of individual variability underlying all of the above effects, despite the smaller variation in implementation quality.

 INSERT TABLE 4 ABOUT HERE

3.4.2. Play relations

The proportion of classmates children identified as those they liked playing with during school breaktimes pre- and post-intervention is shown in Table 4b. As with work relations, there was a general and significant trend towards an increase in numbers over the course of the intervention, $F(1, 477) = 8.13$, $p = .005$, partial $\eta^2 = .02$, but also fluctuations in the actual proportion and degree of change involved. Here, though, none of these was substantial enough to result in significant interaction effects. The extent of change was somewhat weaker than was the case for work relations, suggesting that play relations were affected less directly by the intervention. The somewhat distinct nature of play relations is further borne out by the greater main effect of school area, $F(1, 477) = 30.40$, $p < .001$, partial $\eta^2 = .06$, and the absence of an effect of class type. Despite the confirmation of the expected impact of rural context on play relations in particular, as with work relations higher initial levels did not promote increased growth. This lack of effect could not be attributed to ceiling levels having already been reached in rural classes, as had been anticipated might be the case, since the outcome for these children was comparable change rather than an absence of it. Finally, it should be noted that, once again, the identified effects were subject to considerable individual variation.

3.4.3. Moderating effects of age and gender

Given the widespread individual variability in the pre- and post-intervention indices of social relations, it was pertinent to ask how far this variation was attributable to obvious differences in personal characteristics. In order to ascertain this, correlations were first computed between pupils' age in months and their perceived work and play relations with classmates. Weak positive correlations were found for pre-intervention work and play relations, $r = .24$, $p < .001$, and $r = .11$, $p < .01$, respectively ($n = 526$ children for whom both age and pre-intervention data were available, two-tailed values in both cases), and for change in work relations controlling for pre-intervention score, partial $r = .23$, $df = 474$ based on cases for whom age, pre- and post-intervention data were available, $p < .001$, two-tailed). Age was unrelated, however, to change in play relations. Effects of gender were restricted to play relations, where there was a marginally significant difference between girls and boys pre-intervention (for girls, $M = .38$, $SD = .23$; for boys, $M = .42$, $SD = .20$), $t(537) = 2.00$, $p < .05$, Cohen's $d = .18$; and a significant effect on raw change values, attributable to this pre-intervention difference reducing (for girls, $M = .06$, $SD = .24$; for boys, $M = .01$, $SD = .20$), $t(479) = 2.48$, $p < .05$, Cohen's $d = .23$. Once pre-intervention values were controlled for by taking them as a covariate within a one-way ANCOVA, this effect on change disappeared. In general, then, there were signs that older children gained marginally more from the intervention in terms of work relations, but that with respect to gender, the intervention once again served, if anything, to iron out initial differences.

3.5. Prediction of adjusted change in class relations

To be able to establish how far different aspects of the quality of pupils' experience of group work predicted change in work and play relations, two-level models (student and class) were constructed for each, using MLM techniques. In both cases, the dependent variable was pre- to post-intervention change adjusted for pre-test value, in order to remove the influence of area, class type, age and gender on

baseline ratings, and test for the impact of factors above and beyond these. In line with standard procedures (Luke, 2004), basic intercept and slope models without predictors were computed first, to establish the relative distribution of variance in outcome between student-level and class-level effects, and calculate intraclass correlation coefficients (ICC), that is, the extent to which outcomes for students within the same class were similar. Level 1 models incorporating student variables were then fitted in order to ascertain which parameters had predictive value. Variables considered for inclusion at this stage were: (a) pre-intervention and change values on whichever dependent variable was not the subject of the analysis, to check on cross-influences between indices; (b) total transactive dialogue (propositions and explanations) during the observations of group work in science lessons, as the key student-level predictor of cognitive gains (see Howe et al., 2007)¹; and (c) in the case of work relations, age in months, in view of its significant partial correlation with outcome on this dimension, as reported in Section 3.4.3 above. Class-level indices of the quality of group work, that is, the four S-TOP observational measures and two teacher ratings, were then fitted to generate final models.

3.5.1. Work relations

Table 5 shows the MLM outcomes for change in work relations. The basic intercepts and slopes model identified significant class-level variance around the grand intercept, and the presence of class-level effects was confirmed by the relatively high ICC value of .40. The Level 1 model confirmed further the positive relationship between age and change in work relations, indicating again that older pupils gained more from the intervention than younger ones. It also identified a *negative* effect of transactive dialogue, greater incidence of propositions and explanations acting to suppress growth in work relations. This suggests that whilst dialogue that engenders cognitive conflict and explication of ideas promotes cognitive growth (cf. Howe et al., 2007; Howe & Tolmie, 1998; Piaget, 1932), it can also create tensions in working relationships, perhaps because of its association with disagreement (the proposition/explanation index correlated with disagreements during the same observations at $r = .42$, $n = 130$ children for whom observational data from the group work in science lessons were available, $p < .001$). Both the age and dialogue effects were found to be fixed rather than random, meaning that whilst intercepts varied significantly across classes (i.e., they had differing start points in terms of work relations), the slopes were the same (i.e., the nature of the relationship between the pupil-level predictors and change was constant across classes).

The final overall model (AIC = 76.82, the lowest value among models with related structure; reduction in unexplained class variance = 28.8%, reduction in unexplained pupil variance = 7.7%) modified this picture somewhat. Teachers' rating of the impact of the intervention on their pupils was not significantly related in itself to change in work relations, but interacted with both the age and dialogue effects, the negative estimates indicating that higher levels of perceived child impact were associated with reductions in these effects. Bearing in mind the correlation between child impact ratings and teachers' rating of progress in group-work skills, the positive relationship of CLEF scores to change in work relations helps clarify the processes in operation. The implication is that where group skills were better, work relations

¹ NB as an initial check on direct relations between cognitive and social gains, correlations were computed between standardised change scores on the two measures of science understanding reported in Howe et al. (2007) and adjusted change in work and play relations. None of these correlations was significant.

improved more, and the impact on pupils was rated as more positive. This in turn moderated the benefits of age (suggesting these were partly attributable to older children possessing inherently better group-work skills), and reduced the tensions in work relations produced by greater incidence of transactive dialogue. When CLEF and child impact ratings were included in the model, the negative effect of dialogue was therefore no longer significant at conventional levels.

 INSERT TABLE 5 ABOUT HERE

3.5.2. *Play relations*

The models for change in play relations are shown in Table 6. Compared to work relations, there was substantially less variance at both class and pupil levels (under 60% of the total value there), but the ICC of .49 confirmed nevertheless that there were significant class-level effects in operation. The Level 1 model identified adjusted change in work relations as a fixed positive effect, indicating again consistency across classes in the nature of this effect, and mediation of change in play relations by the impact of the intervention on work relations, in line with other signs that effects on play relations were less direct. The final model (AIC = -29.57, the lowest among those models with similar structure; reduction in unexplained class variance = 27.6%, reduction in unexplained pupil variance = 6.7%) was consistent with this picture. The child impact rating this time exhibited a significant negative influence, and no interaction with other predictors: the greater the perceived positive impact, the lower the change in play relations, perhaps suggesting that teachers focused primarily on apparent productivity in terms of work outcomes. The rating of progress in group skills was a positive influence, though, indicating that play relations again improved more where group skills reached higher levels. Moreover, the effect of change in work relations became non-significant when the child impact and CLEF ratings were included, suggesting its influence was at least in part a function of these broader effects of the intervention, unsurprisingly given the final model for change in work relations.

 INSERT TABLE 6 ABOUT HERE

3.6. *Summary of main findings*

To summarise, then, significant pre- to post-intervention gains were found in both work and play relations, accompanied by the ironing out of initial variations in these associated with class composition and rural versus urban location. Levels of transactive dialogue and group-work skills showed similar improvements and accompanying reductions in variation over the course of the intervention. Multilevel linear modelling identified the primary influence on increased positivity in work relations as improvements in the level of group-work skills as reported by teachers. These also acted to moderate a negative influence associated with the incidence of transactive discussion, despite the positive association of such discussion with achievement. The implication is that productive exchange of ideas created social tensions unless group-work skills became well-developed. Increased positivity in play relations was also found to be related to teacher ratings of the quality of group-work

skills, but the effect was weaker and partially mediated through the change in work relations.

4. Discussion

As noted earlier, Howe et al. (2007) have already reported the achievement gains associated with the present intervention. The demonstration that improvements in work and play relations also took place, verifying Hypothesis 1, therefore confirms that collaborative group work can have a dual cognitive and social impact, at least where it is founded on relational and communicative preparation, in line with the more established effects of cooperative learning. With regard to the social gains, the effect sizes were not large, and individual variation was high. However, these gains were achieved on the basis of a restricted programme that stopped substantially short of wholesale restructuring of classroom activity, and the actual pattern of effects was both clear and internally consistent.

The data also provide for the first time clarification regarding the three potential relationships between cognitive and social gains identified at the outset of this paper. The pre-condition account, at least in any strict version, received little support. Although participants were provided with initial training, they continued to show growth in both group-work skills and transactive dialogue throughout the intervention. The implication is that there was room for improvement over the levels achieved by training, and that these were in that sense sub-optimal. Moreover, it was transactive dialogue during the second and third observations (i.e., whilst it was still increasing) that was predictive of achievement gains (Howe et al., 2007). This would appear to rule out the notion that optimal relational conditions must be in place first, before the cognitive potential of collaborative learning can be realized. Instead, in line with the claims made by Azmitia (2000) and Kreijns and Kirschner (2005), there was clear evidence that students' collaborative skills improve alongside gains in understanding. This is not to deny that initial relational and communicative training may have played a role in the outcomes, but it would seem to have been at most a matter of establishing sufficient minima to permit further growth as part of productive activity.

The lack of effects of social context on either cognitive or social gains points to the same conclusion. Preliminary analyses confirmed that children in the rural schools had a better relational base to draw on at the outset, as anticipated, whilst children in mixed-age classes had better initial work relations. Despite this, Howe et al. (2007) found no difference in achievement gains attributable to these variations. Similarly, in terms of gains in work and play relations, the net effect of the participants' experience of collaborative group work was to iron differences out, rather than to build on them or reinforce them, contrary to Hypothesis 2. This was even true of the initial gender differences in play relations. Instead of well-developed social relations being a pre-condition for effective collaboration, then, it would appear that, provided perhaps certain minimal levels are in place (cf. Blatchford et al., 2006; Kutnick et al., 2008), actual engagement in collaborative learning acts to boost both achievement and social relations to relatively uniform levels, counteracting the effects of social difference. The sole exception to this levelling process was the effect of age, with older children consistently exhibiting better work relations—possibly because of greater time spent in school, and consequently greater experience of working with others. However, even this effect was moderated by improvements in group-work skills during the intervention.

If the pre-condition account can largely be discounted, then the possibility that social benefits are a directly-related consequence of the processes giving rise to achievement gains has little better support, contrary to Hypothesis 3. Certainly, there was no sign that either work or play relations improved as a function of the gains in understanding noted by Howe et al. (2007), or of the mechanisms underlying these. In fact, the key influences on these gains were found to have either no effect on social outcomes (teacher support for transactive dialogue) or else a negative one (transactive dialogue during group work in science lessons).

4.1. Conclusions

The conclusion that social gains are a separate outcome of collaborative learning, in line with Hypothesis 4, therefore seems to be inescapable. However, there would appear nevertheless to be an integral relation between the mechanisms leading to cognitive and social gains, with students' group-work skills forming the critical point of conjunction. Improved work relations were found to be primarily a function of improvements in such skills, this having a knock-on effect on play relations. Improved group-work skills also apparently served to moderate the negative effects of transactive discussion, indicating that it was the experience of successful management of such exchanges that was central to the social benefits. There was no indication that transactive discussion itself actually ceased where group-work skills were lower, threatening achievement gains (cf. the pre-condition account), but there are grounds for thinking that gains in work and play relations may have been absent under these circumstances. Moreover, the convergence over time between levels of group-work skills and transactive discussion suggests that even in terms of achievement gains, group-work skills were facilitatory (cf. Van den Bossche et al., 2005, on the benefits of psychological safety for collaborative learning). In this sense, then, cognitive and social gains would appear to be interlinked, if distinguishable outcomes.

It is important to note that this outline of the processes at work has various points of similarity to the social interdependence account offered in the context of cooperative learning by Roseth et al. (2006), and may help elucidate the exact mechanisms involved there. Certainly, the idea that successful cooperation leads to positive perceptions of co-workers is compatible with the present finding that successful management of collaborative activity led to improved work relations. The fact that this also served to offset the tensions created by transactive exchange is similarly compatible with the argument that positive perceptions of co-workers enhance subsequent cooperation. There would therefore appear to be a real potential for convergent theoretical accounts of the effects of these two types of group work. Further convergence may also be possible between these accounts and work on co-regulation between learners (e.g., Salonen, Vauras, & Efklides, 2005), which emphasizes the role of students' awareness of both their own and others' cognitions and metacognitions in coordinating successful group activity. The present research underscores the possibility that even at elementary school level such awareness is a major strand of what emerges over time from the experience of managing collaborative activity within a stable group context (see also Volet, Summers, & Thurman, 2009), with both greater positive regard and the discounting of tensions stemming from the resultant increase in mutual understanding. This is a point that plainly merits specific investigation.

4.2. Educational implications

The educational implications of the research seem clear. At a basic level, if collaborative group work can achieve social gains alongside gains in understanding, then it is doubly worth teachers investing time and energy in such activities. However, if group-work skills play a central role in facilitating these benefits, then the further implication is that this investment would be repaid to a greater extent if, as argued by Blatchford et al. (2005) and Kutnick and Manson (1998), it is preceded by a period in which preparatory activity is devoted to building up these skills. Given the design of the research reported here, it remains unknown what the outcomes would have been in the absence of such preparation, but there is certainly some reason to think that it provided students with an important head-start, even if its effects were superseded by progress made on the basis of subsequent experience. In terms of quality of dialogue and achievement gains, this point would appear to be confirmed by the outcomes of the similar collaborative learning interventions reported by Blatchford et al. (2006) and Kutnick et al. (2008).

4.3. Limitations of the study

There is much more yet to be investigated in this area, however, and given the relatively low return rate for the initial survey from which participating schools were chosen, it cannot be entirely ruled out that the present sample was in some sense self-selecting or unrepresentative. Moreover, whilst the MLM analyses generated consistent and theoretically illuminating models of the processes involved, they were notably more successful in explaining class-level effects on variation in outcome than individual-level variation. It is plainly the case that the effects of collaborative group work are moderated by a range of other individually varying factors beyond those associated with obvious demographic markers. The nature of these factors stands in need of research if the impact of collaborative group-work programmes is to be optimized. One class-level effect in the present research may perhaps provide something of a start point for such investigation. The reason why children in mixed-age classes had, against expectation, better work relations at the outset than those in single-age classes is unclear, but one possibility is that it was the combined effect of the greater age of some pupils, and thus better inherent group-work skills, plus the willingness of younger members of the class to allow debate to be managed by their older classmates during collaborative activity. The implication is that it is not just group-work skills that matter, but the social dynamics of the context in which they are deployed. It is fine-grained investigation of these interacting effects that seems most likely to repay future investigation.

Acknowledgements

This research was funded by the UK Economic and Social Research Council (ESRC) through its Teaching and Learning Research Programme (TLRP), but they played no role in its conduct or reporting.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, *19*, 716-723.

Aronson, E., Stephan, C., Sikes, J., Blaney, N., & Snapp, M. (1978). The jigsaw classroom. Beverley Hills, CA: Sage.

Arvaja, M., Hakkinen, P., Rasku-Puttonen, H., & Etelapelto, A. (2002). Social processes and knowledge building during small group interaction in a school science project. Scandinavian Journal of Educational Research, *46*, 161-179.

Azmitia, M. (2000). Taking time out from collaboration: opportunities for synthesis and emotion regulation. In R. Joiner, K. Littleton, D. Faulkner, & D. Miell (Eds.), Rethinking collaborative learning (pp. 179-195). London: Free Association Books.

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogue, and the development of scientific reasoning. Social Development, *2*, 202-221.

Bachmann, K., & Grossen, M. (2004). Explanations and modes of collaboration in tutor-tutee interactions at school. In K. Littleton, D. Miell, & D. Faulkner (Eds.), Learning to collaborate, collaborating to learn (pp. 111-132). New York: Nova Science.

Baines, E., Blatchford, P., Kutnick, P., Chowne, A., Ota, C., & Berdondini, L. (2008). Promoting effective groupwork in primary schools: Developing relationships to enhance learning and inclusion. A guide for teachers. London: Routledge.

Bales, R. F. (1950). Interaction process analysis: A method for the study of small groups. Cambridge, MA: Addison-Wesley.

Bell, A., & Sigsworth, A. (1987). The small rural primary school. London: Falmer.

Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., & Chowne, A. (2006). The effect of a new approach to group-work on pupil-pupil and teacher-pupil interaction. Journal of Educational Psychology, *98*, 750-765.

Blatchford, P., Galton, M., Kutnick, P., & Baines, E. (2005). Improving the effectiveness of pupil groups in classrooms. End of award report to the Economic and Social Research Council. Retrieved September 11, 2006 from <http://www.spring-project.org.uk/spring-Publications.htm>

Buchs, C., Butera, F., Mugny, G., & Darnon, C. (2004). Conflict elaboration and cognitive outcomes. Theory into Practice, *43*, 23-30.

Clark, A., Anderson, R.C., Kuo, L., Kim, I., Archodidou, A., & Nguyen-Jahiel, K. (2003). Collaborative reasoning: expanding ways for children to talk and think in school. Educational Psychology Review, *15*, 181-198.

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

Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. International Journal of Educational Research, *58*, 9-19.

Darnon, C., Buchs, C., & Butera, F. (2002). Epistemic and relational conflicts in sharing identical vs. complementary information during cooperative learning. Swiss Journal of Psychology, *61*, 139-151.

Deutsch, M. (1949). An experimental study on the effects of cooperation and competition upon group process. Human Relations, *2*, 199-231.

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In M. Jones (Ed.), Nebraska Symposium on Motivation (pp. 275-319). Lincoln, NE: University of Nebraska Press.

Dillenbourg, P. (1999). What do you mean by 'collaborative learning'? In P. Dillenbourg (Ed.), Collaborative learning: cognitive and computational approaches (pp. 1-19). Oxford, UK: Elsevier.

Doise, W., & Mugny, G. (1984). The social development of the intellect. Oxford, UK: Pergamon.

Edmondson, A. (1999). Psychological safety and learning behaviour in work teams. Administrative Science Quarterly, 44, 350-383.

Foot, H. C., Morgan, M. J., & Shute, R. H. (1990). Children's helping relationships: An overview. In H. C. Foot, M. J. Morgan, & R. J. Shute (Eds.), Children helping children (pp. 3-17). Chichester, UK: Wiley.

Galton, M. J., Hargreaves, L., Comber, C., Wall, D., & Pell, A. (1999). Inside the primary classroom: 20 years on. London: Routledge.

Gillies, R. M., & Ashman, A. F. (2003). An historical review of the use of groups to promote socialization and learning. In R. M. Gillies & A. F. Ashman (Eds.), Co-operative learning: The social and intellectual outcomes of learning in groups (pp. 1-18). London: Routledge Falmer.

Howe, C. J., & Tolmie, A. (1998). Productive interaction in the context of computer-supported collaborative learning in science. In K. Littleton & P. Light (Eds.), Learning with computers: Analysing productive interaction (pp. 24-45). London: Routledge.

Howe, C. J., & Tolmie, A. (2003). Group work in primary school science: Discussion, consensus and guidance from experts. International Journal of Educational Research, 39, 51-72.

Howe, C. J., Tolmie, A., Anderson, A., & Mackenzie, M. (1992). Conceptual knowledge in physics: The role of group interaction in computer-supported teaching. Learning & Instruction, 2, 161-183.

Howe, C.J., Tolmie, A., & Rodgers, C. (1992). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an incline. British Journal of Developmental Psychology, 10, 113-130.

Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., Jessiman, E., & Donaldson, C. (2007). Group work in elementary science: towards organisational principles for supporting pupil learning. Learning and Instruction, 17, 549-563.

Issroff, K. (1995). Investigating computer-supported collaborative learning from an affective perspective. Unpublished PhD thesis, Institute of Educational Technology, The Open University, Milton Keynes, UK.

Jarvelä, S., Lehtinen, E., & Salonen, P. (2000). Socio-emotional orientation as a mediating variable in the teaching-learning interaction: Implications for instructional design. Scandinavian Journal of Educational Research, 44, 293-306.

Johnson, D. W., & Johnson, R. (1979). Conflict in the classroom: Controversy and learning. Review of Educational Research, 49, 51-70.

Johnson, D. W., Johnson, R., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. Review of Educational Research, 53, 5-54.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). The effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin, 89, 47-62.

Kreijns, K., & Kirschner, P. A. (2005, August). Sociable CSCL and CSCoP environments: Warming-up cold CSCL and CSCoP environments. Paper presented at the 11th Biennial Conference of the European Association for Research on Learning and Instruction, Nicosia, Cyprus.

Kutnick, P., & Manson, I. (1998). Social life in the classroom: Towards a relational concept of social skills for use in the classroom. In A. Campbell, & S. Muncer (Eds.), The social child. Hove, UK: Psychology Press.

Kutnick, P., Ota, C., & Berdondini, L. (2008). Improving the effects of group working in classrooms with young school-aged children: Facilitating attainment, interaction and classroom activity. Learning and Instruction, 18, 83-95.

Luke, D. A. (2004). Multilevel modelling. Thousand Oaks, CA: Sage.

MacDonald, R., & Miell, D. (2000). Musical conversations: collaborating with a friend on creative tasks. In R. Joiner, K. Littleton, D. Faulkner, & D. Miell (Eds.), Rethinking collaborative learning (pp. 65-78). London: Free Association Books.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle and high school years. American Educational Research Journal, 37, 153-184.

May, M., & Doob, L. (1937). Cooperation and competition. New York: Social Sciences Research Council.

Mercer, N. (1995). The guided construction of knowledge: Talk amongst teachers and learners. Clevedon, UK: Multilingual Matters.

Piaget, J. (1932). The moral development of the child. London: Routledge & Kegan Paul.

Piaget, J. (1985). The equilibration of cognitive structures. Chicago: University of Chicago Press.

Polvi, S., & Telama, R. (2000). The use of cooperative learning as a social enhancer in physical education. Scandinavian Journal of Educational Research, 44, 105-115.

Raudenbush, S., Bryk, A., & Congdon, R. (2005). HLM 6 – Hierarchical linear and nonlinear modeling. Lincolnwood, IL: Scientific Software International.

Roitashcher, R. C. (1974). A microeconomic model of sociometric choice. Sociometry, 37, 219-238.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), Computer supported collaborative learning (pp. 69-97). Berlin: Springer.

Roseth, C. J., Fang, F., Johnson, D. W., & Johnson, R. T. (2006). Effects of cooperative learning on middle school students: A meta-analysis. San Francisco: American Educational Research Association Convention.

Rourke, L. (2000, May). Operationalizing social interaction in computer conferencing. Paper presented at the 16th Annual Conference of the Canadian Association for Distance Education, Quebec City, Canada.

Salonen, P., Vauras, M., & Efklides, A. (2005). Social interaction: What can it tell us about metacognition and coregulation in learning? European Psychologist, 10, 199-208.

Schwartz, D. L. (1998). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches (pp. 197-218). Oxford, UK: Elsevier.

Shachar, H., & Fischer, S. (2004). Cooperative learning and the achievement of motivation and perceptions of students in 11th grade chemistry classes. Learning and Instruction, 14, 69-87.

Shapira, A., & Madsen, M. C. (1969). Cooperative and competitive behaviour of kibbutz and urban children in Israel. Child Development, 4, 609-617.

Sharan, S. (1980). Cooperative learning in small groups: Recent methods and effects on achievement, attitudes and ethnic relations. Review of Educational Research, 50, 241-271.

Slavin, R. (1989). Cooperative learning and student achievement. In R. E. Slavin (Ed.), School and classroom organization. Englewood Cliffs, NJ: Erlbaum.

Summers, J. J. (2006). Effects of collaborative learning in math on sixth graders' individual goal orientations from a socioconstructivist perspective. The Elementary School Journal, 106, 273-290.

Summers, J. J., & Svinicki, M. D. (2007). Investigating classroom community in higher education. Learning and Individual Differences, 17, 55-67.

Tolmie, A., & Howe, C. J. (1993). Gender and dialogue in secondary school physics. Gender & Education, 5, 191-209.

Tolmie, A., Howe, C. J., Mackenzie, M., & Greer, K. (1993). Task design as an influence on dialogue and learning: Primary school group work with object flotation. Social Development, 2, 183-201.

Tolmie, A., Thomson, J. A., Foot, H. C., Whelan, K., Morrison, S., & McLaren, B. (2005). The effects of adult guidance and peer discussion on the development of children's representations: Evidence from the training of pedestrian skills. British Journal of Psychology, 96, 181-204.

Topping, K. J. (2003). Collaborative Learning Evaluation Form. Dundee: Centre for Peer Learning, University of Dundee. Retrieved August 1, 2005 from <http://www.dundee.ac.uk/fedsoc/CLEF.htm>.

Van den Bossche, P., Van Gennip, N., Gijssels, W., & Segers, M. (2005, August). Collaborative learning as constructive conflict: Building shared mental models. Paper presented at the 11th Biennial Conference of the European Association for Research on Learning and Instruction, Nicosia, Cyprus.

Volet, S., Summers, M., & Thurmann, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? Learning and Instruction, 19, 128-143.

Webb, N., & Farivar, S. (1999). Promoting helping behaviour in cooperative small groups in middle school mathematics. American Educational Research Journal, 31, 369-395.

Wegerif, R. (2000). Applying a dialogical model of reason in the classroom. In R. Joiner, K. Littleton, D. Faulkner, & D. Miell (Eds.), Rethinking collaborative learning (pp. 119-136). London: Free Association Books.

Williams, J. M., & Tolmie, A. (2000). Conceptual change in biology: group interaction and the understanding of inheritance. British Journal of Developmental Psychology, 18, 625-649.

Table 1.
Characteristics of sample

| Class composition | Number of classes | N | Males/ females | Mean age (<u>SD</u>) years, months | Age range | Mean class size |
|-------------------|-------------------|-----|----------------|---|--------------|-----------------|
| Rural | | | | | | |
| Single-age* | 6 | 131 | 60/71 | 10,9 (6.66 months) | 9,3 - 12,4 | 21.8 |
| Mixed-age | 6 | 126 | 67/59 | 10,11 (8.44 months) | 9,1 - 12,5 | 21.0 |
| Urban | | | | | | |
| Single-age | 6 | 156 | 74/82 | 10,11 (7.90 months) | 9,10 - 12,11 | 26.0 |
| Mixed-age | 6 | 162 | 79/83 | 10,8 (7.28 months) | 8,10 - 12,0 | 27.0 |

*Single-age classes were defined as those where students had all started primary education in the same school year, mixed-age as those where students had started in different school years.

Table 2.
Mean proportion of classmates rated pre-intervention as falling into each of four categories of out-of-class relationship (standard deviations in parentheses)

| | Children in rural classes (<u>n</u> = 237) | Children in urban classes (<u>n</u> = 302) |
|-------------------------|---|---|
| Play with out of school | .20 (.17) | .14 (.11) |
| See at club etc | .28 (.28) | .18 (.19) |
| Know family of | .31 (.23) | .17 (.14) |
| See locally | .45 (.34) | .13 (.22) |

Table 3.
Mean scores and distributions on quality of implementation indices

| | <u>M</u> | <u>SD</u> | Range | Skew |
|--|----------|-----------|---------|-------|
| S-TOP Learn (max = 36) | 34.42 | 2.47 | 25 - 36 | -2.69 |
| S-TOP Tasks (max = 63) | 56.50 | 6.45 | 40 - 63 | -1.08 |
| S-TOP Adults (max = 81) | 68.67 | 8.80 | 42 - 81 | -1.28 |
| S-TOP Group Skills (max = 99) | 84.92 | 8.67 | 60 - 99 | -.90 |
| Teachers' ratings of class impact (max = 28) | 24.37 | 2.65 | 17 - 28 | -.96 |
| Teachers' CLEF ratings (max = 34)* | 23.35 | 6.05 | 11 - 34 | -.54 |

n = 24 classes, except * n = 23, due to non-completion by one teacher.

Table 4.

Mean proportion of classmates rated by students pre- and post-intervention as those they a) liked working with, and b) liked playing with during school breaktime (standard deviations in parentheses)

| Class composition | Children in rural mixed-age classes ($\underline{n} = 108$) | Children in rural single-age classes ($\underline{n} = 121$) | Children in urban mixed-age classes ($\underline{n} = 139$) | Children in urban single-age classes ($\underline{n} = 113$) |
|---|--|---|--|---|
| Mean Proportion of Classmates Like to Work With | | | | |
| Pre-intervention | .46 (.32) | .39 (.25) | .47 (.33) | .27 (.24) |
| Post-intervention | .50 (.34) | .47 (.29) | .46 (.29) | .38 (.28) |
| Change | .04 (.27) | .08 (.27) | -.01 (.23) | .11 (.31) |
| Mean Proportion of Classmates Like to Play With | | | | |
| Pre-intervention | .43 (.21) | .48 (.23) | .34 (.20) | .38 (.19) |
| Post-intervention | .45 (.21) | .51 (.26) | .40 (.25) | .38 (.17) |
| Change | .02 (.22) | .03 (.25) | .06 (.22) | .00 (.20) |

Table 5.

MLM outcomes for adjusted change in work relations (proportion class like to work with)

| Parameter | Estimate (B) | p |
|---|--------------|--------|
| Basic intercept and slopes model | | |
| Grand intercept (G_{00}) | .485 | < .001 |
| Class-level variance (U_0) | .041 | < .001 |
| Student-level variance (R) | .061 | |
| Final Level 1 model | | |
| Grand intercept (G_{00}) | .489 | < .001 |
| Age (fixed) (G_{10}) | .013 | .007 |
| Propositions/explanations (fixed) (G_{20}) | -.009 | .014 |
| Class-level variance (U_0) | .033 | < .001 |
| Student-level variance (R) | .058 | |
| Final overall model | | |
| Grand intercept (G_{00}) | .503 | < .001 |
| Child impact (CI) rating (G_{01}) | -.019 | .245 |
| Teacher group skills (CLEF) rating (G_{02}) | .017 | .049 |
| Age (fixed) (G_{10}) | .013 | .008 |
| CI x age (G_{11}) | -.003 | .009 |
| Propositions/explanations (fixed) (G_{20}) | -.006 | .074 |
| CI x propositions/explanations (G_{21}) | -.002 | .039 |
| Class-level variance (U_0) | .029 | < .001 |
| Student-level variance (R) | .056 | |

Table 6.
MLM outcomes for adjusted change in play relations (proportion class like to play with at school
breaktime)

| Parameter | Estimate (B) | p |
|---|--------------|--------|
| Basic intercept and slopes model | | |
| Grand intercept (G_{00}) | .476 | < .001 |
| Class-level variance (U_0) | .029 | < .001 |
| Student-level variance (R) | .030 | |
| Final Level 1 model | | |
| Grand intercept (G_{00}) | .473 | < .001 |
| Adj change in work relations (fixed) (G_{10}) | .199 | .048 |
| Class-level variance (U_0) | .024 | < .001 |
| Student-level variance (R) | .028 | |
| Final overall model | | |
| Grand intercept (G_{00}) | .477 | < .001 |
| Child impact (CI) rating (G_{01}) | -.027 | .048 |
| Teacher group skills (CLEF) rating (G_{02}) | .013 | .018 |
| Adj change in work relations (fixed) (G_{10}) | .186 | .066 |
| Class-level variance (U_0) | .021 | < .001 |
| Student-level variance (R) | .028 | |