ELSEVIER

Learning and Instruction 13 (2003) 569–595

Learning and Instruction

www.elsevier.com/locate/learninstruc

A systematic observational study of teachers' and pupils' behaviour in large and small classes

P. Blatchford *

Institute of Education, University of London, 20 Bedford way, London WC1H OAL, UK

Abstract

Research and debate on class size differences have focused on relations with achievement and there is little relevant research on what mediating classroom processes might be involved. This paper investigates connections between class size and three aspects of teacher and child behaviour in class: teacher–child interactions, pupil attentiveness and off-task behaviour, and peer relations in class. It reports on a systematic observational study of large (average 33 children) and small (average 19 children) reception classes (children aged 4–5 years). Data come from a sub-sample of 235 children in 39 classes drawn from a longitudinal study of two cohorts of over 10,000 children for three years after entry to English infant schools (aged 4–7 years). Results show that teacher–child contacts are more frequent and personalized in small classes, but that children are more likely to be off-task in large classes (whether with the teacher, other children or when not interacting), and interact more extensively with their peers (through work as well as social contacts). The results raise questions about direct models of teacher effects, and suggest that class size is one contextual influence that presents difficulties and opportunities for teachers and children.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Class size; Teacher–pupil interaction; Pupil attentiveness

1. Introduction

There has been a vigorous debate over the educational consequences of class size differences. In the US, the debate has centred on the efficacy and cost effectiveness of class size reduction initiatives, while in the UK the debate has been more about the negative effects of large classes. The most widely quoted research is the experimental

* Tel.: +44-20-7612-6280.
E-mail address: p.blatchford@ioe.ac.uk (P. Blatchford).

0959-4752/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0959-4752(02)00043-9

Tennessee STAR project. Although findings are still contentious (see Grissmer, 1999), there is some agreement that results show that smaller classes, at least below 20, have positive effects on pupil academic performance (e.g. Finn & Achilles, 1999). There is also general agreement that the benefits of smaller classes are most pronounced if introduced immediately after school entry, that is, with the youngest children in school (Blatchford & Mortimore, 1994).

However, the STAR project, like many others, was predominantly interested in the relationships between class size and academic outcomes, and had little to say about classroom processes that might explain effects found. This is recognized by Finn and Achilles (1999, p. 102), two original members of the STAR research team, who argue that: "Despite dozens of earlier studies, the classroom processes that distinguish small from large classes have proven elusive." In summing up recent US research and debate on class size reductions, Grissmer (1999) has concluded that there is a lack of coherent theories by which to guide and interpret empirical work, and with which to make new predictions. The situation in the UK is much worse, in the sense that there is little research to guide debate on class size effects, and such research as does exist is limited in terms of research methods (Goldstein & Blatchford, 1998).

We need, therefore, accounts of classroom behaviour that might explain why smaller classes make a difference and benefit academic achievements. These accounts would need to show ways in which classroom processes are altered as a result of class size differences, and in this paper we examine one set of possible mediating variables. It might be expected that a central classroom process affected by class size would be teachers' and pupils' behaviour and this is the focus of the paper. Results are drawn from a sub-sample of classes from a large-scale longitudinal and 'correlational' study based at the Institute of Education, University of London. The aim of the study, overall, was to provide a full analysis of class size difference effects on educational attainment, taking into account the full range of class sizes (not just large and small), as well as detailed information on pupil characteristics, class compositional factors, other classroom process variables including within class grouping practices and curriculum experiences, teacher characteristics (such as age and experience) (see Blatchford, Baines, Kutnick, & Martin, 2001; Blatchford, Edmonds, & Martin, in press; Blatchford, Goldstein, Martin, & Browne, 2002; Blatchford, Moriarty, Edmonds, & Martin, 2002). In this paper we explore in detail differences between a sub-set of large and small classes in just one type of classroom process (i.e. teacher and pupil behaviour). Three aspects of teacher and pupil behaviour are considered in this study.

1.1. Teaching behaviour

Research on teaching has a long and varied history, and is far too numerous to be reviewed here. Reviews of this research show that there are different and often conflicting paradigms of research on teaching (Shulman, 1986), but a central tenet of many studies is the importance of maximizing teaching time and instructional support for children's learning. This is expressed most obviously in the process-

product tradition which has stressed the importance of maximizing students' academically engaged time in classrooms, and the important influence of teacher instructional time and active teaching (Brophy & Good, 1986; Creemers, 1994). This has been supported by more recent research on effective teaching, more allied to effective schools research (see Creemers, 1994; Galton, Hargreaves, & Pell, 1996). A quite different approach to teaching, with roots more in cognitive psychology and post-Vygotskian approaches, and which stresses notions like scaffolding and contingent learning environments (see Tharp & Gallimore, 1991; Wood, 1998; Wood & Wood, 1996), also gives a central place to maximizing adult instructional behaviour.

A main point to make, and one which informs this paper, is that despite the voluminous literature of teaching, research evidence on the effect of contextual influences such as size of class is not consistent or well developed. In Dunkin and Biddle's (1974) model, which supported research in the 'process-product' tradition, there were four stages—presage, context, process and product. Class size, in this model, is one of several context variables. Bennett (1996) used this model in his report of a Teacher Association survey of teachers', parents', and school governors' views on class size effects. This model serves to position class size in the context of overall influences on children's progress, but there has been little research which has studied connections between class size and the 'processes' within classrooms, which will include teacher and pupil behaviour.

On logical and common sense grounds it seems likely that the number of children in a class will increase the amount of time that teachers spend in procedural and domestic matters such as taking the registers, lining children up and putting on coats, dealing with domestic duties such as toileting, accidents, etc. (e.g. Bassey, 1996) and, conversely, decrease the amount of time that can be spent on instruction and dealing with individual children. It might be expected to be particularly important to maximize the amount of teaching and individual support for the youngest children in school, who are least likely to have developed independent strategies of learning.

This expectation is consistent with teachers' views. In the UK, Bennett's (1996) account of a survey of teachers' and headteachers' views shows that practitioners believe large class sizes affect teaching and learning, and they were particularly aware that larger classes could have an adverse effect on the amount of teacher attention. Pate-Bain and others report, on the basis of teacher interviews conducted at the end of each school year in the STAR research:

> A common benefit cited by teachers in small and regular plus aide classes was that they were better able to individualize instruction. These teachers reported increased monitoring of student behaviour and learning, opportunities for more immediate and more individualized teaching, more enrichment, more frequent interactions with each child, a better match between each child's ability and the instructional opportunities provided, a more detailed knowledge of each child's needs as a learner, and more time to meet individual learners' needs using a variety of instructional approaches (Pate-Bain, Achilles, Boyd-Zaharias, & McKenna, 1992, p. 254).

Teachers' reports are supported by the meta-analysis conducted by Glass, Cahen, Smith, and Filby (1982), who found that smaller classes resulted in greater teacher knowledge of pupils, frequency of one-to-one contacts between teachers and pupils, variety of activities, adaptation of teaching to individual pupils, and opportunities to talk to parents (cf. Cooper, 1989). Other studies report more individual teaching and attention (Harder, 1990; Pate-Bain et al., 1992; Turner, 1990), and more feedback (Cooper, 1989; Pate-Bain et al., 1992).

However, in direct contrast to teacher views, Shapson, Wright, Eason, and Fitzgerald (1980), in a widely cited study, found no statistically significant differences between class sizes for most teacher activities, subject emphasis, classroom atmosphere or quality measures, and teachers did not alter the proportion of time spent interacting with the whole class, with groups or with individuals. One of the few observation measures on which there were differences was the proportion of pupils addressed as individuals. This increased in a linear way as class sizes decreased from 37, 30, 23, to 16. Worryingly, they found that these observation results were at odds with teachers' own views. There was, in other ways, a gap between professional judgement and observational research findings.

There have been several more recent studies that have examined the effects of class size on teaching. Molnar et al. (1999) report results from the Wisconsin Student Achievement Guarantee in Education (SAGE) project—a 5-year K-grade three project begun in the 1996–1997 school year. Though not a study of class size reduction as such, the programme required participating schools to implement four interventions, one of which involved pupil teacher ratio reductions to 15 students per teacher. Teachers were asked to rank items in terms of the extent to which they were affected by reduced class size. The teacher behaviours that received the highest rankings were more individualized instruction; more teaching time; more discussion, sharing and answering; more hands-on activities and more content coverage. The most important classroom process, affected by reduced class size, according to Molnar et al. (1999), is therefore individualization.

In another recent study, Betts and Shkolnik (1999) modelled relationships between class size and teacher time allocation, on the basis of a secondary analysis of a national survey of students in middle and high schools in the USA (the survey covered the period 1987–1992). Data came from a survey of maths teachers who completed forms describing retrospective estimates of minutes per week in group instruction (it is not clear if this includes whole class teaching), individual instruction per student, and percentage time in instructional activities. They provide some evidence that there is a tendency for teachers to substitute group instruction for individual instruction as class size increases, and devote less time to group instruction and more on individual instruction in smaller classes. There was small effect on percentage overall instructional time. They argue teachers would make better use of small classes if they did not reduce group instruction.

Rice (1999) also conducted a secondary analysis of teacher survey data. Data came from a national panel survey of students transferring to high. Measures are aggregates at the class level of high school maths and science teachers' estimates of percentage time devoted to instructing small groups and individuals, time devoted to innovative

instruction, time devoted to whole group discussion, minutes of homework assigned, time devoted to administrative tasks, and time devoted to maintaining order in high school mathematics and science classes. Findings showed that in maths as class size increased less time was spent on small groups and individuals, innovative instructional practices, and whole group discussions, though increases in class sizes beyond 20 had little effect. There were no relationships between class size and instructional time allocation measures in science.

Both these last two studies raise interesting questions about the effect of class size differences on teaching and non-teaching (e.g. procedural/managerial) time overall, and also on how it is shared between individuals, groups and the whole class. But the studies are limited in that they rely on a secondary analysis of rather general teacher retrospective estimates of time spent, and basic distinctions, for example, between individual, group and class contexts, which might be expected to be differently affected by class size, are not always clear. Another problem with the studies is the age of students involved. We have seen that greatest effects are reported with younger children, and especially children immediately after entry to school (Blatchford & Mortimore, 1994), and so results involving much older children may underestimate effects of class size differences and/or involve different processes.

1.2. Pupil attentiveness and off-task behaviour

One main element of children's successful adjustment to school is likely to involve their attentiveness in class, e.g. as reflected in the extent of their attention to the teacher and to work, or their 'on task' behaviour, as it has been commonly called. The theoretical basis for this is well known, and usual references are to Carroll and Bloom. As we have seen, a central notion of research in the process-product tradition was the importance of maximizing students' academically engaged time in classrooms (Creemers, 1994; Pellegrini & Blatchford, 2000). Regardless of any connection with class size, studies of classroom processes related to pupil achievement, going back decades, support the view that a key aspect of educational success is attentiveness, or active learning time, time on task or some equivalent term (e.g. Creemers, 1994). There is a good deal of evidence that pupil inattentiveness is a major variable having negative effects on pupils' achievement (Rowe, 1995).

A main focus of this paper is on the connection between class size and pupils' on task and off-task behaviour. Common sense and logic suggest that with more children in the class there will be more potential for distraction, and more possibility of being off-task. Conversely, in small classes there will be more opportunities to engage children and keep them on task. This might be expected to be particularly relevant in the case of the youngest children in school who are less likely to have developed the capacity for independent work.

Blatchford and Mortimore (1994) in their review found several studies which indicated that pupils in smaller classes attend more and spend more time on task (Carter, 1984 and Cahen in Cooper, 1989; Klein, 1985). They also participate more (Cahen in Cooper, 1989), and are more absorbed in what they are doing (Cahen in Cooper, 1989). Cahen and others (1983 in Cooper, 1989) argue that pupil attention is greater

in smaller classes because pupils are not lost in the crowd and have more opportunities for participating. The authors speculate that the effect of class size on attentiveness is most pronounced in the case of low attainers, because teachers can bring them out more.

Finn and Achilles (1999, p. 103) have expressed the connection between small classes and pupil attention or engagement in class perhaps most clearly. They argue that:

> The evidence indicates that the key to the benefits of small classes is increased student engagement in learning. In a small class, every student is in the firing line. It is difficult or impossible to withdraw from teaching–learning interactions in a small-class setting. Social psychologists have long recognized the negative relationship between group size and participation of individuals—the principle underlying concepts such as 'social loafing' and 'diffusion of responsibility'…When class sizes are reduced, the pressure is increased for each student to participate in learning, and every student becomes more salient to the teacher. As a result, there is more instructional contact, and student learning behaviours are improved.

In the previous section we looked at relationships between class size and teaching. It might also be expected that the amount of opportunity for pupils to engage in interactions with their teachers, which would include interactions about school tasks, would be affected by size of class. In one observational, longitudinal study of children (Tizard, Blatchford, Burke, Farquhar, & Plewis, 1988), it was found that children's active interactions with teachers (times when they either initiated or responded to teachers) decreased over the first three years of schooling. It might be expected that in a larger class there would be fewer opportunities for such active involvement in interactions with teachers, but, as in the case of teaching time, research is not clear cut. It must also be said that Shapson et al. (1980), in a systematic observation study, did not find that pupils in smaller classes participated more in assigned tasks.

1.3. Peer relations in class

There is a lot of evidence that children's early social behaviour toward peers is an important predictor of later social and personal adjustment (Parker & Asher, 1987). The effects of children's aggressive, withdrawn and prosocial behaviours toward peers have received most empirical support. There is also a large literature on the value of collaborative or cooperative group work in classrooms (see review in Galton, 1990), and naturalistic studies of children's interactions in classrooms, which will include those with peers (e.g. Galton, Simon, & Croll, 1980; Tizard et al., 1988). However, as in the case of teaching and pupil on task behaviour, there is little research on the effects of contextual classroom factors like class size on peer relations. Research on children at nursery level indicates that less favourable staff pupil ratios can lead to more negative relations between children, including more aggression, annoying and teasing. This research has also found that increasing the

numbers of adults can lead to more talking between adults rather than to pupils (Smith, McMillan, Kennedy, & Ratcliffe, 1989). Smith and Connolly (1980) found that there were higher levels of aggression when there is more overcrowding in preschool settings. But other research with older pupils seems less clear, and Shapson et al. (1980) found no difference in conflicts between pupils.

As with teaching, therefore, associations between peer relations and contextual factors like size of class are not clear. One might expect that in large classes children will be more distracted and that this would include more off-task behaviour with their peers. This distraction might take an externalizing form in the sense of overtly disruptive behaviours and 'mucking about', or a more internalizing form in the sense of being disengaged and distracted from work. One might also expect more negative and aggressive behaviours between children in larger classes, and for teachers to more easily miss squabbles between children.

1.4. Summary of aims of the paper

This paper reports on research, which examines connections between class size differences and three dimensions: teaching behaviour, pupil attention/off-task behaviour in class, and peer relations in class. While previous research and theory does not allow exact predictions, several expectations informed collection and analysis of data:

(a) in the case of teachers in larger classes there would be less teaching time overall, fewer one-to-one contacts between teacher and child, fewer interactions within which a pupil was the focus of a teacher's attention, and more procedural/routine interactions;
(b) in the case of pupils in larger classes there would be more off-task and less on task behaviour, less evidence of active contributions to classroom interaction with teachers in terms of initiations and responses, less attention to a teacher and more activities related to procedural/routine matters;
(c) with regard to peer relations, in large classes there would be more disruptive, disengaged and negative interactions between children.

1.5. Methods used in research on class size differences

There are two main features of the approach adopted in this study, which it is appropriate to comment on at this point. The first concerns the overall research strategy. Elsewhere we have reviewed research methods used in studies of class size effects (Blatchford, Goldstein, & Mortimore, 1998; Goldstein & Blatchford, 1998), and have identified limitations with much early UK research, particularly the use of cross-sectional designs which make interpretation of relationships between class size and outcomes problematic. It is often assumed that experimental designs, and randomized controlled trials more specifically, are superior, and necessary to provide the basis for causal interpretations. However, as we have argued elsewhere (Goldstein & Blatchford, 1998), randomized designs are questionable, theoretically in terms of the

validity and generalizability of results, and also in terms of their usefulness for policy recommendations. It is difficult to generalize from studies, like the STAR project, in which teachers (and pupils) are assigned to classes of a limited range of different sizes, and problems of validity are particularly likely in cases where experimental assignment only lasts for a short time (see Galton et al., 1996). There are also immense practical and financial difficulties in setting up large-scale studies of class size differences.

The overall strategy in the present study has been to employ a longitudinal research design, random selection of participating schools, and measurement of effects of naturally occurring differences in class size using measures carefully drawn up on the basis of previous research and pilot work. A naturalistic design is more useful in addressing policy issues in that it is more 'authentic', and reflects adjustments and processes as they occur under normal circumstances. The study had two aims, overall; it was set up to examine connections between: (a) size of class and pupils' progress, and (b) size of class and classroom processes, such as teacher and pupil behaviour, within class grouping practices, teacher self-perceptions, assessment and record keeping. The study followed two large cohorts of pupils. Because the first year of school was of particular interest, and because the observation component was very labour-intensive, in this paper we report just on a sub-sample of reception classes (i.e. children aged 4–5 years) from one cohort (Cohort 1). The general strategy, therefore, which we explain in detail below, has been, for this systematic observation component, to select classes which differed in terms of whether they were small (under 20) or large (over 30). In this way, classes differed in terms of class size but this was naturally not artificially created.

A second general feature concerns the method used to collect data on teacher and pupil behaviour. As we have seen, there are a number of limitations concerning data collection in previous research that make interpretation and summary of research findings difficult. Methods used in studies are not always clearly described or adequate. Much is relatively anecdotal and based on teacher report and the reported experience of individual teachers. Though valuable, there are questions about the validity and generalizability of such views, especially given Shapson et al.'s (1980) findings of discrepancies between teacher reports and classroom observation data. Large-scale secondary analyses have involved relatively crude, if easily quantified, retrospective judgements of time spent. Methods used in observation studies can be problematic; there are, for example, reliability problems with data drawn from post hoc analyses of narrative observation records.

One way to measure more reliably the extent to which class size differences are related to the type and nature of teacher–pupil interactions and pupil work related behaviour is through the use of systematic observation. This collects data in terms of categories previously defined on conceptual grounds and refined in pilot work. In contrast to other forms of data collection it builds up data on the basis of careful recording of on-going behaviour (rather than, say, ratings or judgements). Criticisms of systematic observation have usually centred on validity issues (e.g. Delamont & Hamilton, 1986), but it can be a useful research tool when answering specific research questions where data are needed on relatively easily observed, high fre-

quency behaviours (Croll, 1986; McIntyre & Macleod, 1986). This was the method used in Shapson et al. (1980) study which, though widely cited, is dated now, used an experimental design, as questioned above, and was based in one country (Canada), with accompanying questions about generalizability to other countries.

The aim was to address precisely the three sets of predictions given above on the basis of a systematic observation study making use of classroom observation techniques adapted from previous research at the Institute of Education (Blatchford, Burke, Farquhar, Plewis, & Tizard, 1987; Tizard et al., 1988). The category system used was at heart a simple 3×4 matrix with 12 cells. One dimension denoted three 'social modes'—whether a child interacted with the teacher, other children or the child was on its own—and the other dimension referred to whether the content of the behaviour was academic, procedural, social or off-task.

2. Method

2.1. Sample: schools, classes and children

The overall Class Size Project (from which the observation sample was drawn) followed for three years a large cohort of pupils who entered reception (aged 4–5 years) classes during 1996–1997, and a second separate cohort of pupils who entered reception classes one year later during 1997–1998. More than 220 schools were involved in the 1996–1997 cohort, with more than 300 classes and 8000 children in eight Local Education Authorities (LEAs). A further five LEAs joined the Project in 1997–1998. The children were followed for the first three years: reception, Year 1 and Year 2 (4–7 years). The research design involved random selection of schools within the participating LEAs. All children entering reception classes in a selected school during the year were included in the study. There were a number of different kinds of data collected in the study, including start of school and end of year assessments, termly teacher completed questionnaires, teacher and headteacher completed end of year questionnaires, pupil behaviour ratings (PBR) on each child in the study, and systematic observations and case studies conducted on sub-samples of the main sample. This paper concentrates on results from the systematic observations.

Schools for the observation study were selected from the total sample. Three of the participating LEAs were approached and agreed to take part in the observation component. Schools were selected on the basis of information already provided on class sizes. Schools with small (20 or under) and large (30 and over) reception classes (i.e. classes for children on entry to school, who became 5 years during the year) were identified and a random selection approach was made to see if they were willing to take part. The aim was to get 40 classes, divided between large and small classes. In the event 39 classes in 27 schools with the required characteristics agreed to take part. There were 18 large and 21 small classes. Those identified as small classes had on an average 19.4 children and those identified as large classes had 32.5 children on the register, according to the observer notes at the time of the observations. For each class, observers were provided with the names of six children—three boys and

three girls—randomly chosen by the researchers, along with two reserves to be observed in cases where the sample children were absent. In the event there were observations on 235 children (one class had observations on seven children).

2.2. Data collection

It is clearly important to establish that there are no prior differences between children in large and small classes that might affect results from the observation study. The relationship between class size and background factors such as income level and attainment level is not clear-cut in the UK but in order to check if there were prior differences between large and small classes, three sources of data were used:

(a) School entry attainment: information was collected when the pupil entered school by means of a baseline entry assessment conducted by the teacher. The procedure was the Avon Reception Entry Assessment (1996), which covers literacy and mathematics and comprises information from teacher ratings, based on classroom observations, and tasks completed by children. A measure of literacy knowledge was derived by adding for each child scores on 15 items in language, 18 in reading, 17 in writing and a test of letter identification (how many of 26 letters were recognized in terms of either name or sound), and a measure of mathematics was based on total correct out of 19 items. Training was provided for class teachers in its use.
(b) Information on free school meal entitlement was collected for each child.
(c) The PBR Scale was an adaptation of Ladd and Profilet's (1996) child behaviour scale and is a 58 item teacher administered rating instrument with each item marked on a three point scale ('certainly applies to this child', 'applies sometimes to this child', 'does not apply to this child'). The main factors are—hyperactive/distractable (15 items), aggressive (14 items), anxious/fearful (three items), prosocial (seven items), asocial (seven items), and excluded (seven items). The hyperactive/distractible scale was used to assess the degree of externalizing behaviour problems for children in classes. Cronbach's alpha for this scale was high (0.96).

2.2.1. Observation procedures

There has been much debate about the most appropriate sampling technique for systematic observation, in particular about whether to code behaviour continuously as events occur in 'real time' or whether to use one of several forms of time sampling techniques (e.g. Sackett, 1978). Ideally, continuous coding is preferable because it can capture precise information on frequencies and duration, but in practice it can become too difficult with a relatively large number of behaviours and/or where behaviour occurs relatively frequently. Time sampling techniques were introduced as long ago as the 1930s and 1940s (see Arrington, 1943) to overcome this problem.

A systematic observation schedule developed in previous research (Blatchford et al., 1987; Tizard et al., 1988) was used. This involved direct, that is to say, on the

spot, observations of selected children in terms of previously developed categories and in terms of 5-min observation sheets divided into continuous 10-s time samples. The schedule was child-based in the sense that one child at a time was observed, the 'target' child. The aim was to provide a description of the child's behaviour; teachers and other children were observed only when they came into contact with the target when he/she was being observed.

The schedule had categories that provided a description of time spent in different settings (e.g. teacher directed activities), different school subject areas (e.g. language and mathematics) and a description of how children behaved when with their teachers, other children and when not interacting (the three social modes). Within each of these three 'modes' were categories that covered work, procedural, social and off-task activity.

Observation categories and brief definitions and examples are as follows:

- Setting—*Individual work*: the child is working on his/her own 'work activities' (i.e. reading, writing and number, or any other activity leading to a product, e.g. painting) and is not directed by the teacher. *Groupwork*: a group, involving the target child but not the teacher, is working together on work activities as defined under 'individual work' (above). *Play*: 'free' play activities not resulting in a 'product', for example, sand, water, pretend play. *Teacher-led situations*: that is to say, where the target child is involved. *Transition times*: for example, tidying up, lining up, etc.
- Subject—*Maths*: mathematical activities specifically set up by the teacher (e.g. weighing, sets, sorting, measuring, worksheets) as well as apparatus/activities with a clear mathematical focus (e.g. number games). *Language*: includes reading activities (e.g. silent reading, reading to teacher, flashcard work), writing activities (writing 'news' or stories, copying teacher's writing) and stories, discussions of points of language (e.g. letter sounds), 'news', etc. *Other taskwork*: non-language or maths activities with some end product (e.g. construction materials, painting) or clearly defined procedures (e.g. instructional card games). 'Topic', nature, science, history and geography would be included here. *Any other activity*: for example, sand and water play, dressing up, transitions.
- Teacher to Child Contact—*Social setting*: *one-to-one*, *group* or *whole class*. *Child role*: *focus* (target child is focus of teacher's attention) or *audience* (another child is focus in group or class involving target child, or teacher interacts to same extent with all children). These two sets of categories described the behaviour coded in the 'teacher content' section. *Teacher content*: *task-teach*: contacts directly concerned with the substantive content of children's task activities, i.e. communicating concepts, facts or ideas by explaining, informing, demonstrating, questioning, suggesting. *Task-preparation*: contacts directly concerning the organization and preparation of children's task activities and not their substantive content. *Task-silent*: a teacher's contribution to task contact is passive, e.g. hearing child read, looking over child's work. *Procedure*: contacts concerned with classroom management and organization of classroom routine, often at transition times, e.g. milk, washing, changing, organizing materials. *Social*: personal or social comments, e.g.

about life outside the classroom, children's appearance, health, etc. *Unclear*: not possible to code reliably.

- Child to Teacher Contact—*Child contribution*: codes child's contribution to interaction with teacher in terms of *respond* to teacher, *initiate* contact with teacher, *attend* to teacher, *continued* interaction from previous time intervals and *unclear*. These categories describe the child's contribution to the behaviour coded in the 'child content' section. If the child interacted in an overt way ('respond', 'initiate', 'continued'), these were coded. Only when the child attended for the whole 10-s interval was 'attend' coded. Because of its likely low frequency of occurrence, initiate was given priority over respond if both occurred within the same interval. Predominant activity sampling, therefore (see below) was not used for the 'child contribution' categories. *Child content*: *task*: all child behaviours in contact with teacher that are concerned with 'task' as defined for teacher content, above. *Procedure*: equivalent to teacher 'procedure', above. *Social*: equivalent to teacher 'social', above. *Inappropriate*: child behaviour to teacher obviously unrelated to teacher request or situation, for example, not answering a question on maths, but making a comment about a television programme the previous evening. *Off-task*: child behaviour involving the teacher, but not directed at her, that is, inappropriate or unrelated to situation (e.g. not attending to story). *Unclear*: behaviour from child to child that cannot be coded reliably, as above.
- Child to Child Contact—Coded when child is in contact with other children but not teacher. *Task*: all contacts with other children that are concerned with the content of 'tasks' as defined for teacher content, above. *Procedure*: all contacts with other children concerning classroom organization and routine. *Social*: social or personal contacts not related to work or procedure. *Mucking about*: contacts that involve fooling around. Like social contacts, they are not about task or procedural activities, but are more obviously off-task. *Aggressive*: target child is aggressive (verbally or physically) towards other child(ren). *Help*: target child helps another child, for example, helps tie her shoelaces. *Unclear*: behaviour with other children that cannot be coded reliably, as above.
- Not Interacting—Coded during time intervals child is not in contact with teacher or other children. *Task involved*: target child is involved in own 'task' activity (as defined for teacher content, above). *Procedure*: activity concerned with procedure or routine. *Off-task* (*active*): target child focuses on something other than task in hand. *Off-task* (*passive*): target child is disengaged during task activity, for example, wandering around or daydreaming. *Audience*: target child observes other children or teacher when not in contact with them. *Unclear*: behaviour when not interacting that cannot be reliably coded, as above.
- Conventions for coding categories—Target children's behaviour was coded for 5-min periods divided into 30 consecutive 10-s time intervals. Within intervals, choices within sets of mutually exclusive behaviours (i.e. 'setting', 'subject', 'social setting', 'child role', teacher content, child content, 'child–child', 'not interacting') were made on the basis of which occurred for the longest time, that is, the predominant activity in each interval. Behaviour was coded only once if

it occurred within an interval. A stopwatch was used to indicate beginnings and ends of time intervals, and when 5 min had elapsed.

In predominant activity sampling only that behaviour from a mutually exclusive set which occurs for the longest time within a set is coded. This is an accurate method of sampling to the extent that behaviours within a set have an equal chance of occurrence. Problems of interpretation arise if there is bias against certain categories; for example, behaviours that always occur briefly within a time interval. It is difficult to assess this, but there are a number of reasons that lead us to believe that it was not serious.

First, using a relatively short time interval can minimize the choices required between behaviours in a set. We have no test of this, but it is probable that in the majority of intervals there was no choice because only one behaviour in a set occurred; for example, children tend to stay in maths, teacher-led situations, individual work, and so on, in blocks of time and therefore throughout a 10-s interval. Secondly, having a number of different sets, by which one simultaneously coded behaviour within an interval, can help to minimize the choices between categories. For example, teacher–child interaction within an interval may have been coded in terms of 'setting', 'subject', social setting, child role, teacher content, child contribution, and child content.

2.2.2. Organization of observations

Observations were not conducted during parts of the day when Physical Education, music, swimming, assembly, milk, etc. took place. Usually these were times when children left the classroom, sometimes to be with another teacher. It was difficult to work to an exact timetable, such is the informality and flexibility of most infant classrooms—milk time, for example, may be merely an accompaniment to a teaching situation. The basic principle was to observe during classroom-based work activities, i.e. those parts of the day when language, maths, other work like craft and painting, and free play in the classroom could have taken place.

The aim was to observe the six children in each class five times per day, that is, 25 min per child and two and half hours observation per school day, and 15 observation sheets per child overall. Given that there were parts of the school day, such as assembly, music and movement, etc. when observations were not possible, this was probably the maximum possible. It could vary to some extent, depending on events in class on that day. The aim was to observe each child for three days in each class. Each of the six children was observed in turn. Inevitably there were absences from school, and in order to complete observations on each child it was sometimes necessary to make extra return visits.

In the event the average number of completed observation sheets per child was 14, and there were 3238 sheets overall. This amounted to 97,140 10-s observations overall (30 per sheet), and there were on an average 413 of these observations per child (this number fell short of the theoretical maximum—450—because of missing data, observer error and codes not being readable). The data presented in tables are therefore the average number of 10 s time intervals in which a particular behaviour

category was coded. (Total frequencies of teacher to child (or child to teacher), child–child and not interacting behaviours did not always add up to 413 because of missing data.) In terms of time, there were 69 min of observation per child, which amounted to 270 h for the whole sample. Observations were conducted over a period of a few weeks at the same time during the Spring term. Observations within classes were as far as possible on consecutive days, though this was not sometimes possible because of events in schools, that kept children out of class, e.g. rehearsals in the hall, and child absences requiring revisits.

2.2.3. Observers

The observers were recently retired senior teachers and headteachers, contacted through participating LEAs. They were experienced in education, knew their way around schools, and, it was hoped, able to put teachers and pupils at their ease. The basic aim was to get them to avoid passing judgments, and to use the schedule as intended. It was stressed that the accuracy and reliability of the results depended on this. All observers had a day's initial training in which they were provided with an observation manual of categories, conventions and procedures, as well as tips acquired during previous use. Conventions were discussed and there was work on videotapes, accompanied by periodic checks of accuracy and understanding of how to use categories. This was followed by at least a day's observation in a reception class not involved in the study, and then a follow up training session to discuss field visits and iron out difficulties.

2.2.4. Reliability checks

In a previous work using the schedule, extensive reliability checks were carried out (see Blatchford et al., 1987). After the training period, reliability checks were carried out by synchronized but independent coding of the same events by 'trainee' observers and the research officer who had main responsibility for devising the schedule (so-called criterion coding—see Frick & Semmel, 1978). Matrices were then drawn up showing, for each set of categories, how trainee and 'expert' coded each time interval.

It is preferable to use agreement measures that utilize agreements and disagreements on an event-by-event basis (nominal agreement) rather than comparison of total category frequencies (marginal agreement) because it uses more information and is a more stringent test (Frick & Semmel, 1978). For this reason observer agreement was calculated by Cohen's kappa, which gives a measure of nominal agreement after inflation by chance agreement has been subtracted. This means that, although observers are likely to agree that a rare behaviour did not occur, the value of kappa allows for this (see Plewis & Bax, 1982). Kappa is best used to give a measure of agreement for a set of at least two categories, and so agreement measures are in the main for sets of categories, e.g. teacher to child content, child–child content, not interacting, etc. Results were calculated separately for each trainee–expert pairing. As few differences were found between these pairings a combined matrix was drawn up and kappa recalculated. Reliability coefficients for the main sets of mutually exclusive categories were high. Setting, subject, teacher–child social setting, child

role, teacher content, child to teacher, child contribution, child content and not interacting all had reliability coefficients (kappa) greater than 0.80. Kappa for child–child content was 0.77 (Blatchford et al., 1987).

2.2.5. Statistical analysis

All the results that follow are based on total numbers of 10 s observations in observation categories per child. In some tables, percentages of all 10-s observations or percentages within each broad category (e.g. teacher–child) are also given. The statistical significance of differences between large and small classes was tested with analysis of variance with the child's score as the unit of analysis.

3. Results

3.1. Differences between children in large and small classes

In order to ensure that there were no existing differences between children in large and small classes, that might affect interpretation of observation results, differences between classes in terms of (1) school entry assessments in literacy and mathematics, (2) eligibility for free school meals (the percentage in a class eligible for free meals), and (3) externalizing behaviour problems (average scores for each class) were analysed with ANOVA. There were no differences between classes on these three measures, indicating that children did not differ on entry to small and large classes in terms of attainment, poverty/income levels, or behaviour problems.

3.2. The three social modes: interactions with teachers, other children and behaviour when not interacting

Total scores for the three social mode categories were calculated for each child by adding each of the teacher to child content categories (and the child to teacher content categories, which are almost but not exactly the same, see below), the child–child categories, and the not interacting categories. These total scores therefore give a broad picture of how children's time was distributed. Mean differences between large and small classes in terms of these variables are shown in Table 1.

Table 1
Systematic observations of large and small classes: mean differences in the social modes

		Mean (s.d.)		F (df) significance
		Small (n = 122)	Large (n = 122)	
1	Teacher–child			
	(a) Teacher to child	212.95 (104.37)	143.93 (56.89)	38.47 (1,232), $p < 0.001$
	(b) Child to teacher	215.25 (103.77)	148.01 (58.16)	36.50 (1,232), $p < 0.001$
2	Child–child	53.96 (41.15)	76.36 (48.38)	14.64 (1,232), $p < 0.001$
3	Not interacting	130.80 (73.30)	154.10 (65.07)	6.56 (1,232), $p < 0.05$

Note: n = number of children.

It can be seen in Table 1 that children in small classes were more often observed interacting with their teachers than were children in large classes. The teacher to child and child to teacher totals can differ slightly; for example, a child might initiate a contact in a time interval but the teacher has not yet or does not respond to it. Just taking the teacher to child measures, in a small class children were observed on an average 213 times with their teacher, as compared to 144 times in a large class.

Looking more closely at the teacher to child categories, it can be seen in Table 2 that children in smaller classes were more likely to interact with their teachers on a one-to-one basis (36 vs 18 observation), and in groups (77 vs 35 observations). There were no differences between large and small classes in the amount of whole class teaching. The child was more likely to be the 'focus' of a teacher's attention (173 vs 117 observations), that is to say, the child was the subject of a teacher's attention on a one-to-one basis (by definition the child was the focus), or in a group or the whole class.

Children in small classes also experienced more teaching overall. Just taking the category 'task teach', which might be considered the most strict definition of teaching, in the sense that it denotes contacts directly concerned with the substantive content of children's task activities (communicating concepts, facts or ideas by explaining, informing, demonstrating, questioning, etc.), rather than setting them up (task preparation) or getting materials ready, etc. (procedure), we see that this was coded on an average 156 times per child in small classes but 101 times in large classes. There was also about twice as much teacher 'task preparation' in small as opposed to large classes. There was more teacher social talk with children in small classes, for example, talk about their clothes, home experiences, etc. though this

Table 2
Systematic observations of large and small classes: mean differences in teacher to child categories

	Mean (s.d.)		*F* significance	Percentage of sub-set
	Small (n = 122)	Large (n = 112)		
Social setting				
One-to-one	36.07 (25.70)	17.50 (17.85)	40.28 ($p < 0.001$)	14.2
Group	76.95 (84.88)	35.04 (30.28)	24.44 ($p < 0.001$)	29.8
Class	110.60 (54.97)	101.26 (62.73)	n.s.	55.9
Child role				
Focus of teacher's attention	173.24 (100.42)	116.69 (58.24)	27.13 ($p < 0.001$)	
Content				
Task teach	155.61 (90.78)	100.51 (47.93)	32.84 ($p < 0.001$)	70.2
Task preparation	28.97 (24.82)	15.51 (15.22)	24.48 ($p < 0.001$)	12.7
Task silent	9.54 (14.02)	3.68 (7.72)	15.31 ($p < 0.001$)	4.8
Procedure	13.08 (12.58)	21.52 (20.95)	14.21 ($p < 0.001$)	7.2
Social	5.75 (10.52)	2.71 (6.46)	6.95 ($p < 0.01$)	3.0
Unclear	2.30 (7.70)	4.08 (8.22)	n.s.	2.2

Note: ANOVAs df (1,232).

category was not frequent overall. The only category of teacher talk to children that was more numerous in larger classes was procedural talk, as predicted.

The categories involved in children's contacts with their teachers are given in Table 3. The first set of categories covers the type of interaction with the teacher, e.g. attending, initiating or responding. There was more attending to the teacher (144 vs 96), responding (27 vs 22) and more initiating (8 vs 6, though this category was not common) in small classes. So children had a more active role in interaction with their teachers (respond plus initiate), and were more attentive to her (attend). There are also signs that interactions are more continuous or sustained in small classes (i.e. continued from one interval to the next).

Table 3 also shows results covering the content of the behaviour. As with the teacher to child categories, there is more task related talk (166 vs 107) and more social talk to teachers in small classes. Conversely, in large classes there was more inappropriate or off-task behaviour (usually not attending to the teacher). There was also more child to teacher talk about procedural matters in smaller classes, which is surprising given the above results showing there was more teacher to child procedural talk in large classes.

Table 4 shows the categories involving interactions with other children. About 60% of the child–child contacts are classified as task—that is to say, concerned with allocated work—and there were more of these in large classes. This proportion is in line with the greater amount of child–child contacts overall in large classes. However, there were also more off-task behaviours, i.e. more inappropriate and off-task behaviours in larger classes, though these were not common overall—in about 5% of all child–child contacts, they were mucking about. But, still, on an average there

Table 3
Systematic observations of large and small classes: mean differences in child to teacher categories

	Mean (s.d.)		*F* significance	Percentage of sub-set
	Small (n = 122)	Large (n = 112)		
Child contribution				
Respond	26.61 (17.61)	21.62 (13.31)	5.92 ($p < 0.05$)	11.7
Initiate	8.25 (7.52)	5.95 (6.02)	6.59 ($p < 0.05$)	3.3
Attend	144.39 (78.31)	95.61 (47.68)	32.41 ($p < 0.001$)	69.6
Continued	29.19 (34.04)	8.30 (12.29)	37.65 ($p < 0.001$)	13.6
Unclear	2.25 (5.85)	3.14 (4.95)	n.s.	1.8
Child content				
Task	166.40 (89.67)	106.90 (45.23)	39.97 ($p < 0.001$)	78.8
Procedure	21.48 (17.54)	15.00 (13.49)	9.92 ($p < 0.01$)	9.3
Social	7.08 (14.05)	3.27 (7.49)	6.54 ($p < 0.05$)	3.8
Inappropriate	0.94 (2.66)	1.44 (4.25)	n.s.	0.3
Off-task	10.28 (17.23)	15.01 (20.03)	3.77 ($p = 0.05$)	5.4
Unclear	2.46 (5.78)	5.39 (9.34)	8.49 ($p < 0.01$)	2.4

Note: ANOVAs df (1,232).

Table 4
Systematic observations of small and large classes: mean differences in child–child interactions

	Mean (s.d.)		F	p	Percentage of sub-set
	Small (n = 122)	Large (n = 112)			
Task	34.02 (29.28)	45.49 (36.43)	7.10	$p < 0.01$	57.6
Procedure	7.21 (8.41)	6.47 (7.58)	n.s.	n.s.	11.1
Social	9.42 (11.57)	15.22 (14.27)	11.76	$p = 0.001$	21.8
Mucking about	2.13 (3.74)	5.39 (7.70)	17.41	$p < 0.001$	5.4
Aggressive	0.10 (0.39)	0.19 (0.99)	n.s.	n.s.	
Help	0.20 (0.75)	0.80 (2.32)	7.53	$p < 0.01$	0.9
Unclear	0.89 (1.87)	2.79 (5.15)	14.70	$p < 0.001$	3.1
Total	53.96 (41.15)	76.36 (48.38)	14.64	$p < 0.001$	

were double the number of off-task contacts between children in large classes, and this seems more than just reflecting more time overall with other children. There were also more social contacts between children in large classes. There were no differences between class sizes in the amount of interactions involving procedural matters. In larger classes there are also more social contacts between children (9 vs 15). There were very few coded instances of aggression and help between children. So in summary there are more contacts overall between children in large classes, involving both task and off-task behaviours.

Table 5 shows categories under the not interacting mode. Again children in large classes were more likely to be involved in their work—task involved—and again the proportion reflected the greater amount of time they spend not interacting. But again, as with child–child interactions, children in large classes are also more likely to be off-task. There were two forms of this, either active or passive. The first is actively doing something other than the allocated work and not acceptable to the teacher, while the second—and more frequent—is behaviour that is disengaged from

Table 5
Systematic observations of small and large classes: mean differences in not interacting categories

	Mean (s.d.)		F	p	Percentage of sub-set
	Small (n = 122)	Large (n = 112)			
Task involved	88.87 (53.04)	108.22 (58.39)	7.06	$p < 0.01$	75.2
Procedure	17.53 (16.57)	12.73 (10.75)	6.78	$p = 0.01$	11.6
Off-task (active)	2.98 (6.54)	6.42 (9.19)	11.04	$p = 0.001$	2.4
Off-task (passive)	4.44 (6.71)	13.07 (13.93)	37.41	$p < 0.001$	6.5
Audience	16.57 (19.34)	11.72 (12.90)	5.00	$p < 0.05$	4.0
Unclear	0.41 (1.66)	1.93 (3.87)	15.68	$p < 0.001$	0.3
Total	130.80 (73.30)	154.10 (65.07)	6.56	$p < 0.05$	

the work. Together these amounted to about 9% of all not interacting categories, and there is much more in larger classes. There were three times as much of the passive disengaged form of off-task behaviour in large classes.

3.3. Total off-task behaviour

We can add up all the off-task behaviours in the three social modes to give a total off- task score for each child (i.e. the total of child to teacher inappropriate and off-task, child to child mucking about and aggressive, and not interacting off-task active and off-task passive—see Table 6). Social activities we have excluded here because they were by definition not deemed to be unacceptable to the teacher. A more precise term for 'off-task', as used here, might be something like 'task avoidant', though we retain the more common usage here.

It can be seen in Table 6 that there is twice as much off-task behaviour overall in large classes in comparison to small classes (42 vs 21 observations). The most frequent forms of off-task behaviour are not attending to the teacher and not attending to their work when on their own.

There was then consistently more off-task behaviour in large classes across all three social modes. What about the other main summed categories of behaviour, i.e. on task, procedure and social? In contrast to off-task behaviour, on task behaviour was affected by social mode; as we have seen, there was more child to teacher task related behaviour in small classes but more child to child and not interacting task related behaviour in large classes. There was more procedure related behaviour overall in small classes, and as we have seen, this was explained by more child to teacher and procedure when not interacting in small classes.

3.4. School subject and setting

We now turn to results concerning school subject categories. Overall (i.e. across small and large classes), most time was spent in language activities (mean 136

Table 6
Systematic observation of small and large classes: mean differences in child off-task behaviours

	Small	Large	F	p	Overall off-task behaviours (%)
Child–teacher					
Inappropriate	0.94 (2.66)	1.44 (4.25)	n.s.	n.s.	4
Off-task	10.28 (17.23)	15.01 (20.03)	3.77	$p = 0.05$	41
Child–Child					
Mucking about	2.13 (3.74)	5.39 (7.70)	17.41	$p < 0.001$	12
Aggressive	0.10 (0.39)	0.19 (0.99)	n.s.	n.s.	–
Not interacting					
Off-task (active)	2.98 (6.54)	6.42 (9.19)	11.04	$p = 0.001$	15
Off-task (passive)	4.44 (6.71)	13.07 (13.93)	37.41	$p < 0.001$	28
Total	20.87 (25.00)	41.52 (36.01)	26.32	$p < 0.001$	100

observations), followed by other activities (e.g. sand and water play, dressing up; 111 observations), other task work (83 observations), and mathematics (68 observations). The most noticeable difference between small and large classes was in the case of language activities ($F(1,232)22.68$, $p < 0.001$); there were also significantly more maths activities ($F(1,232)4.82$, $p < 0.05$). However, there was more 'other task work' in large classes ($F(1,232)10.61$, $p < 0.001$). There was no difference between large and small classes in 'other activities'. This means that children in small classes experienced more basic language and maths activities, while in large classes there was more of the other activities such as topic work, history and geography. With regard to the setting categories, as would be expected, there were more teacher-led activities in small classes ($F(1,232)7.04$, $p < 0.01$) and conversely, more play activities in large classes ($F(2,232)7.22$, $p < 0.01$).

4. Discussion

As was shown earlier, research and debate on class size differences have focused on relations with achievement. The position taken in this paper is that too little is known about classroom processes, affected by class size differences, which might mediate any relationship with outcomes. In this paper, connections between class size and one set of classroom processes—involving teacher' and pupils' behaviour—have been explored.

4.1. Teaching behaviours

The effect of class size on teaching behaviour seemed clear. There was consistent evidence, as predicted, that children in small classes were more likely to interact with their teachers, there was more teaching on a one-to-one basis, more times when children were the focus of a teacher's attention, more teaching overall, and more times when children were attending to the teacher and actively involved in interactions with them (i.e. responding or initiating rather than just attending). These results therefore show that individual children in small classes receive more interactions with their teachers of a task related nature. It is not possible to say with present evidence much about the quality of these interactions, but it seems safe to conclude that there is more likelihood of individualized teaching in small classes. In general, these results appear consistent with other studies reviewed earlier, for example, Molnar et al. (1999) evaluation of the SAGE initiative. Although there is more one-to-one teaching in small classes, the greater incidence of times when the child was the focus of attention indicates that children receive more attention in group and whole class situations as well. This means we need to be clear about the nature of individualization affected by small compared to large classes; the support received by children applies to all social contexts, whether individual, group or whole class contexts. This needs to be borne in mind when considering worries that smaller classes might encourage an over reliance on individual teaching.

The results indicate that children in small classes have a more active involvement

in interactions with their teachers, as indexed by the greater number of initiations and responses. The trend toward individualization in small classes is not therefore indicative of a passive role for children; the opposite seems more likely; that is to say, children in large classes spend less time actively interacting.

One result, concerning the child's role in interaction with their teachers, was at variance with the teacher's contact with children. While they engaged in a greater amount of procedural interactions in small classes, teachers directed more procedural interactions to children in large classes. It may be that this just reflects the greater likelihood of child to teacher contacts in small classes. In other words, in small classes children interact more with their teachers, and this is evidenced in all categories of behaviour, including contacts involving procedural matters. Children in large classes are less likely to contact their teachers in all ways, whether this is related to the task, social matters or procedural matters. However, from a teacher's perspective, the situation may be different. In small classes they can concentrate on task and social matters, but in large classes they may be forced in the interests of classroom management to spend more time getting children ready, showing them where to find materials, etc. In smaller classes teachers can spend more time on task related matters and less on procedural.

Results concerning school subjects also show that children in small classes spend more time in the basic areas of language (particularly) and mathematical activities, and less in other aspects of the curriculum like geography, history or topic work. This may be explained in a similar way to the results already discussed. In other words, teachers with fewer children can maximize basic task related activities, while in a larger class there may be more procedural activities.

Children also receive more interactions from their teachers of a social nature in small classes, indicating that the interactions are more personalized. The fact that there are fewer teacher interactions about procedural matters adds to this picture. Overall in smaller classes children seem to experience interactions that are more work and socially intense. Turning this on its head, in a large class a child will, by comparison, experience a less intense contact with his/her teachers, receiving fewer work and social contacts but more contacts in group and whole class contexts about procedural matters, as well as spending less time on the basic subjects of language and maths.

The present results are informative about the balance of individual, group and whole class contexts for teacher–pupil interactions. In contrast to Betts and Shkolnik (1999), interactions in groups did not decrease with size of class; rather, teaching in both individual and group contexts increased. It seems that teaching overall increases as class size decreases, and this was reflected in the greater amount of teaching in individual and group contexts. It might be argued that one solution to the teacher's difficulties in contacting children in large classes would be to alter her approach so that there is more teaching to larger groups or the whole class. But there was no evidence that teaching to the whole class increased in larger classes, and this ran contrary to expectation. This might suggest that teachers in large classes strive to maintain the same balance of individual, group and whole class teaching as their colleagues in small classes. There is a probable age of child effect, in that

teachers of such young children are unlikely to feel comfortable about increasing the amount of whole class teaching. This may be possible in some curriculum areas, and for some activities, but will inevitably be of limited relevance to primary aged children, especially the youngest children, who may be more easily distracted.

What are the implications of the present results for theoretical positions referred to earlier? The present study did not set out to exactly test a particular theoretical position, but at a general level, with regard to primary schools, there is support for an interpretation of class size effects in terms of teacher time allocation. They show that more time is spent in teaching and in literacy and mathematics. As explained in more detail by Blatchford et al. (2002), this suggests a kind of 'dilution' theory, as seen for example, in research on family size. Meadows (1996) explains family size effects on children's cognitive development in terms of the parents' effect being diluted if they spread themselves between many children. This term might be applied to the allocation of a teachers' time to children in large classes, i.e. attention to individual children becomes more diluted across more children.

The present results can be seen in terms of research on teaching, introduced earlier. As Galton et al. (1996) have pointed out, small classes can encourage aspects of teaching that are the same as those identified in research on effective teaching, linked with the promotion of pupil achievement, going back decades. Again, the connection will not necessarily follow, and small classes will not necessarily make a bad teacher better, but small classes seem likely to make it easier for teachers to be effective. These results to some extent overlap with dimensions usually identified as constituting effective tutoring. Meadows (1996), for example, reviews post-Vygotskian theories and studies to show that a number of dimensions seem to characterize effective tutoring. These include individual attention, immediacy of feedback, sustained interactions, flexible and effective questioning techniques, and so on. These are all instances of what is commonly called 'scaffolding' (see Wood, 1998, for origins of this term). Not all the dimensions are supported by the present study; but there is a strong suggestion that in a small class a teacher will more easily be able to provide at least some aspects of effective scaffolding for her pupils. As Meadows points out, and consistent with research on class size effects (see Blatchford & Mortimore, 1994), scaffolding is most important in the early years of schooling, when it needs to be at its most active and sustained.

In terms of an approach to instruction favouring scaffolding, school classrooms seem inherently disadvantaged in terms of providing a 'contingent learning environment' (Tharp & Gallimore, 1991; Wood & Wood, 1996), because teachers cannot provide the individualization and personalization of instructional interaction needed for all children in the class. This problem is likely to be magnified with a large number of children. In these terms, small classes can help minimize the problem.

4.2. Pupil on task and off-task behaviour

Turning to the child's role in classrooms, children in large classes were, as predicted, more likely to be off-task—on present evidence twice as much time, compared to small classes—and this was true of all three social modes. Children were

less likely to attend to the teacher and to be off-task in contacts with her, more likely to be actively off-task with other children, and more likely to be off-task when on their own, especially in the passive form of being disengaged from allocated work. These results appear to offer support for Finn and Achilles'(1999) view that class size affects children's engagement in class.

It should be noted that the results on off-task behaviour, just considered, support the connection between class size and a *lack* of engagement in class. The situation with regard to on task behaviour is intriguing and instructive. (Results concerning class size effects on task behaviour are not just the opposite of off-task behaviour—i.e. as on task behaviour increases, on task behaviour necessarily decreases—because other behavioural categories could also be coded within a time interval, e.g. social and procedural.) The present results indicate that the frequency of on task behaviour is conditional on the social mode. When interacting with their teachers, children are more likely to be on task in small classes; however, children in larger classes spend more time on task in contact with other children and when on their own, i.e. when not interacting. This situation reflects the fact that in large classes children, overall, spent more time with other children and on their own and, accordingly, less time with their teachers. As on task behaviour is the main form of behaviour within each social mode it follows the same pattern.

It is not true to say, therefore, that in larger classes there is simply less on task behaviour; overall there is less on task behaviour, but this masks the fact that in larger classes there can be more on task behaviour with other children and when on their own. This raises questions about learning and interactions in different social modes, which are themselves affected by size of class. If one starts from the assumption that teacher–child contacts are likely to be the most conducive social context for learning and achievement then the situation is worrying. However, one should not quickly dismiss the view that task related contacts with peers are important. On present evidence we cannot say much about quality of peer task interactions. Previous research has shown that interactions can be relatively at a low level (Bennett, Desforges, Cockburn, & Wilkinson, 1984; Galton, 1990). But the present results show that there are more of all types of child–child interaction categories in larger classes, which will include social contacts. So just as teacher–child contacts are more extensive and personalized in smaller classes, it may be that peer interactions are more extensive and personalized in larger classes, and more intense. In short, class size affects the balance between teacher–child and child–child relationships.

In this paper, differences between large and small classes have been found in a number of aspects of teacher and pupil behaviour. It has been shown that these differences in behaviour are likely to be attributable to class size differences and are unlikely to be caused by any prior differences between children in large and small classes, in that there were no differences on entry to school in terms of academic assessments of literacy and maths, and poverty or income levels, or in terms of teachers' ratings of behavioural difficulties.

The purpose of this paper was specifically to examine in detail whether small and large classes affected one form of classroom process, that is, teacher and pupil behaviour. This aim affected the sample size and selection. One needs, therefore, to be

cautious about drawing conclusions concerning the role of behavioural measures in relation to attainment. Connections between class size and other classroom processes have been found in other components of the overall research programme, and are reported elsewhere. These include within class grouping practices (Blatchford et al., 2001), other aspects of teaching (Blatchford et al., 2002), curriculum experiences, including individual reading support (Blatchford, in press), and teachers' professional satisfaction and stress (Moriarty, Edmonds, Blatchford, & Martin, 2001). A full modelling of influences on progress in academic achievement is beyond the remit of this paper, but would need to take these, and other, processes into account.

One kind of qualification needs to be attached to these results and conclusions. We have seen that children of this age, when placed in large and small classes, differ in terms of opportunities for more teaching, one-to-one and individualized support, but these conclusions can only strictly be applied to other classes with the same aged children. As described above, it was decided to concentrate on the first year of school, because previous research and educational theory suggest that this is most likely to be affected by class size differences (Blatchford & Mortimore, 1994). There are concerns about growing class sizes in later stages in primary schools in the UK over Key Stage 2 (7–11 years), and in current research the author is studying the same children over KS2 to see if class size effects are still evident.

Finally, the results need to be seen in the context of conceptual models of the influence of class size on pupil achievement. In this paper we have looked at relationships between class size and one form of classroom process, that is to say, teacher and pupil behaviour. The results raise questions about traditional ways of viewing teaching effects. Classroom processes have tended to be viewed in terms of teachers' actions toward pupils and pupils' learning or attainments. There is an underlying assumption, in other words, of a direct model, with teaching having a causal effect on pupils' learning. But teachers do not meet pupils out of context, and class size can be seen as one contextual influence on classroom life that plays a part in the nature of the interactions between teachers and pupils. The conceptual roots of this view can be found in Bronfenbrenner (1979) and the ecological psychology approach. The basic idea was that the classroom context has distinctive forces or 'signals', different to other contexts, which pull events and participants along with them (Kounin & Gump, 1974). Different class sizes may well have different forces or signals. Bearing this contextual approach in mind, it is instructive to look at those views, which have downplayed the importance of class size differences. Rivkin, Hanushek, and Kain (2000) have recently argued that teaching quality and training are important, and that class size is not. A British Government agency report (OFSTED, 1995) dismissed class size in favour of teacher quality. The rather convenient conclusion that follows from this argument is that blame is attributed to the teacher and the model is direct; that is to say, teacher to pupils. But the present study suggests, in contrast to the direct model, that it is not entirely the teacher's responsibility. Class size is one environmental contextual factor that will influence teachers and pupils in a number of ways. The evidence in this paper indicates ways that a small class, while not making a poor teacher good, can allow teachers to provide opportunities for more extensive and task related individualization of instruc-

tion. Conversely, larger classes can present almost inevitable difficulties and hard choices for teachers, at least in terms of their interactions with children. A teacher may well have less time for each child to have a turn, less time to contact children whether in groups or individually. It is not, therefore, as Rivkin et al. (2000) have implied, a case of either supporting teacher training to improve teacher quality, or reducing class sizes. We need to consider both together, and ways of making the most of the opportunities provided by smaller classes.

References

Arrington, R. E. (1943). Time sampling in studies of social behaviour: a critical review of techniques and results with research suggestions. *Psychological Bulletin*, *40*(2), 81–124.

Avon Reception Entry Assessment (1996). Avon Education Department and the Institute of Education.

Bassey, M. (1996). Inspection, unlike research, is disinterested! *Research Intelligence*, *55*, 30–32.

Bennett, N. (1996). Class size in primary schools: perceptions of head teachers, chairs of governors, teachers and parents. *British Educational Research Journal*, *22*(1), 33–55.

Bennett, N., Desforges, C., Cockburn, A., & Wilkinson, B. (1984). *The quality of pupil learning experiences*. London: Lawrence Erlbaum.

Betts, J. R., & Shkolnik, J. L. (1999). The behavioural effects of variations in class size: the case of math teachers. *Educational Evaluation and Policy Analysis*, *21*(2), 193–213.

Blatchford, P. *The class size debate: Is small better?* Milton Keynes: Open University Press, in press.

Blatchford, P., Baines, E., Kutnick, P., & Martin, C. (2001). Classroom contexts: connections between class size and within class grouping. *British Journal of Educational Psychology*, *71*, 283–302.

Blatchford, P., Burke, J., Farquhar, C., Plewis, I., & Tizard, B. (1987). An observational study of children's behaviour at infant school. *Research Papers in Education*, *2*(1), 47–62.

Blatchford, P., Edmonds, S., & Martin, C. Class size, pupil attentiveness and peer relations. *British Journal of Educational Psychology*, in press.

Blatchford, P., Goldstein, H., Martin, C., & Browne, W. (2002). A study of class size effects in English school reception classes. *British Educational Research Journal*, *28*(2), 169–185.

Blatchford, P., Goldstein, H., & Mortimore, P. (1998). Research on class size effects: a critique of methods and a way forward. *International Journal of Educational Research*, *29*, 691–710.

Blatchford, P., Moriarty, V., Edmonds, S., & Martin, C. (2002). Relationships between class size and teaching: a multi-method analysis of English infant schools. *American Educational Research Journal*, *39*(1), 101–132.

Blatchford, P., & Mortimore, P. (1994). The Issue of Class Size for Young Children in Schools: what can we learn from research? *Oxford Review of Education*, *20*(4), 411–428.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Brophy, J., & Good, T. (1986). Teacher behaviour and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching*. New York: Macmillan.

Cooper, H. M. (1989). Does reducing student-to-teacher ratios affect achievement? *Educational Psychologist*, *24*(1), 79–98.

Croll, P. (1986). *Systematic classroom observation*. London: Falmer.

Creemers, B. (1994). *The effective classroom*. London: Cassell.

Delamont, S., & Hamilton, D. (1986). Revisiting classroom research: a continuing cautionary tale. In M. Hammersley (Ed.), *Controversies in classroom research*. Milton Keynes: Open University Press.

Dunkin, M. J., & Biddle, B. (1974). *The study of teaching*. New York: Holt, Reinhart and Winston.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, *21*(2), 97–109.

Frick, T., & Semmel, M. I. (1978). Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research*, *48*(1), 157–184.

Galton, M. (1990). Grouping and group work. In C. Rogers, & P. Kutnick (Eds.), *The social psychology of the primary school*. London: Routledge.

Galton, M., Hargreaves, L., & Pell, A. (1996). *Class size, teaching and pupil achievement*. Report for National Union of Teachers. Leicester University.

Galton, M., Simon, B., & Croll, P. (1980). *Inside the primary classroom*. London: Routledge & Kegan Paul.

Glass, G., Cahen, L., Smith, M. L., & Filby, N. (1982). *School class size*. Beverley Hills, CA: Sage.

Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: a review of methodology with particular reference to study design. *British Educational Research Journal*, *24*(3), 255–268.

Grissmer, D. (1999). Class size effects: assessing the evidence, its policy implications and future research agenda. *Educational Evaluation and Policy Analysis*, *21*(2), 231–248.

Harder, H. (1990). A critical look at reduced class size. *Contemporary Education*, *62*(1), 28–30.

Klein, K. (1985). The research on class size. *Phi Delta Kappan*, *66*, 578–580.

Kounin, J. S., & Gump, P. V. (1974). Signal systems of lesson settings and the task-related behaviour of pre-school children. *Journal of Educational Psychology*, *66*(4), 554–562.

Ladd, G. W., & Profilet, S. M. (1996). *Child behavor scale*. IL: University of Illinois at Urbana-Champaign.

McIntyre, D., & Macleod, G. (1986). The characteristics and uses of systematic observation. In M. Hammersley (Ed.), *Controversies in classroom research*. Milton Keynes: Open University Press.

Meadows, S. (1996). *Parenting behaviour and children's cognitive development*. Hove, East Sussex: Psychology Press.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: a pilot program in targeted pupil–teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, *21*(2), 165–177.

Moriarty, V., Edmonds, S., Blatchford, P., & Martin, C. (2001). Teaching young children: perceived satisfaction and stress. *Educational Research*, *43*(1), 33–46.

OFSTED (1995). *Class size and the quality of education*. London: OFSTED.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: are low-accepted children at risk? *Psychological Bulletin*, *102*(3), 357–389.

Pate-Bain, H., Achilles, C. M., Boyd-Zaharias, J., & McKenna, B. (1992). Class size makes a difference. *Phi Delta Kappan*, *74*(3), 253–256.

Pellegrini, A., & Blatchford, P. (2000). *Children's interactions at school: peers and teachers*. London: Edward Arnold.

Plewis, I., & Bax, M. (1982). The uses and abuses of reliability measures in developmental medicine. *Developmental Medicine and Child Neurology*, *24*, 388–390.

Rice, J. K. (1999). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. *Educational Evaluation and Policy Analysis*, *21*(2), 215–229.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2000). *Teachers, schools and achievement*. Paper to Amherst College. University of Rochester.

Rowe, K. J. (1995). Factors affecting students' progress in reading: key findings from a longitudinal study. *Literacy, Teaching and Learning*, *1*(2), 57–110.

Sackett, G.P. (Ed.). (1978). *Observing behaviour, Vol. II. Data collection and analysis methods*. Baltimore, MD: University Park Press.

Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, *17*, 144–152.

Shulman, L. S. (1986). Paradigms and research programs—the study of teaching: a contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Smith, A. B., McMillan, B. W., Kennedy, S., & Ratcliffe, B. (1989). The effect of improving pre-school teacher/child ratios: an 'experiment in nature'. *Early Child Development and Care*, *41*, 123–138.

Smith, P. K., & Connolly, K. (1980). *The ecology of preschool behaviour*. London: Cambridge University Press.

Tharp, R., & Gallimore, R. (1991). A theory of teaching as assisted performance. In P. Light, S. Sheldon, & M. Woodhead (Eds.), *Learning to think*. London: Routledge.

Tizard, B., Blatchford, P., Burke, J., Farquhar, C., & Plewis, I. (1988). *Young children at school in the inner city*. Hove: LEA.
Turner, C. M. (1990). Prime time: a reflection. *Contemporary Education*, *62*(1), 24–27.
Wood, D. (1998). *How children think and learn* (2nd ed.). Oxford: Blackwell.
Wood, D., & Wood, H. (1996). Vygotsky, tutoring and learning. *Oxford Review of Education*, *22*(1), 5–16.